

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 09:10:56 ; Search time 6478 Seconds
(without alignments)
11278.571 Million cell updates/sec

Title: US-09-508-658A-3

Perfect score: 1545
Sequence: 1 agagaagtgaggctcttc.....aaaataataaattagctg 1545

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 200 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1541.8	99.8	1545	9	AB006683 Homo sapi
2	1540.2	99.7	1545	6	CQ731863 Sequence
3	1367.8	88.5	1463	9	AB006685 Homo sapi
4	1023.6	66.3	2027	9	AB006682 Homo sapi
5	1023.6	66.3	2245	6	A97284 Sequence 1
6	1023.6	66.3	2245	6	HSAPBCHD
7	1023.6	66.3	2258	6	CQ731864 Sequence
8	462.4	29.9	20000	9	AB006684 Homo sapi
9	462.4	29.9	36284	9	HSAP9610 Homo sapi
10	462.4	29.9	42133	9	AP001060 Homo sapi
11	462.4	29.9	110000	2	AC003656_5
12	462.4	29.9	340000	9	AP001754 Homo sapi
13	451.2	25.0	143192	9	AL954247 Pan trogl
14	385.6	25.0	1656	10	AF128772 Mus muscu
15	385.6	25.0	1656	10	AF128773 Mus muscu
16	385.6	25.0	1659	6	A97293 Sequence 10
17	385.6	25.0	1659	10	MMU132243 Mus muscu
18	385.6	25.0	1906	10	MMU24821 Mus muscu
19	385.6	25.0	1924	10	AF128116 Mus muscu

20	385.6	25.0	1936	10	AF079536 Mus muscu
21	375.6	24.3	1921	10	AF128117 Mus muscu
22	375.6	24.3	1933	10	AF128115 Mus muscu
23	336	21.7	2586	9	HSJLTPK
24	174.6	11.3	1744	10	AF128121 Mus muscu
25	174.6	11.3	1747	10	AF128120 Mus muscu
26	174.6	11.3	1756	10	AF128119 Mus muscu
27	174.6	11.3	1759	10	AF128118 Mus muscu
28	159	10.3	1625	10	AF128124 Mus muscu
29	159	10.3	1637	10	AF128122 Mus muscu
30	149	9.6	1622	10	AF128125 Mus muscu
31	149	9.6	1634	10	AF128123 Mus muscu
32	118.4	7.7	9876	6	AX344994 Sequence
33	101.8	6.6	9876	6	AX344995 Sequence
34	96.6	6.3	581	11	BV089377 RPAMSEQ0
35	96.6	6.3	18351	10	AF105002 Ratius no
36	96.6	6.3	18616	10	MMU007715 Mus muscu
37	96.6	6.3	146872	10	AF073797 Mus muscu
38	96.6	6.3	158049	2	AC015891 Mus muscu
39	96.6	6.3	190019	2	AC138672 Mus muscu
40	90.2	5.8	217688	2	AC108592 Ratius no
41	90.2	5.8	283951	2	AC109383 Ratius no
42	82.6	5.3	558	11	BV163805 RPAMSEQ0
43	82.6	5.3	574	11	BV089350 RPAMSEQ0
44	82.6	5.3	593	11	BV088992 RPAMSEQ0
45	77.2	5.0	125020	9	AF429315 Homo sapi
46	73.6	4.8	125020	9	AF429315 Homo sapi
47	70	4.5	184594	5	BX649390 Zebrafish
48	65	4.2	9466	9	AK127046 Homo sapi
49	65	4.2	9466	9	AF425231 Homo sapi
50	63.2	4.1	6438	10	BC058578 Mus muscu
51	63.2	4.1	67573	2	AC122760 Mus muscu
52	63.2	4.1	173909	2	AC134529 Mus muscu
53	62.4	4.1	572	11	BV089378 RPAMSEQ0
54	61.4	4.0	248835	2	AC097256 Ratius no
55	61.4	4.0	273874	2	AC127615 Ratius no
56	60.4	3.9	121375	5	BX004850 Zebrafish
57	60	3.9	60	6	CQ541728 Sequence
58	60	3.9	6327	6	A56817 Sequence 1
59	60	3.9	6328	6	AR224019 Sequence
60	60	3.9	6328	6	AR269336 Sequence
61	60	3.9	6328	6	AR67436 Sequence
62	60	3.9	6412	6	CQ722409 Sequence
63	60	3.9	6417	6	AX332238 Sequence
64	60	3.9	6417	6	HSMT2218 Sequence
65	60	3.9	6475	6	AR338834 Sequence
66	60	3.9	6584	6	BC038596 Homo sapi
67	60	3.9	132592	2	AL589792 Homo sapi
68	60	3.9	166518	9	HS120622 Human DNA
69	59.8	3.9	1053	6	AR383186 Sequence
70	59.8	3.9	1591	9	HSU08379 Human M1-2
71	59.8	3.9	6331	6	CQ491372 Sequence
72	59.8	3.9	6331	6	AX924075 Sequence
73	59.8	3.9	6331	6	AX924076 Sequence
74	59.8	3.9	6331	9	AP006515 Homo sapi
75	59.8	3.9	6771	6	AX24049 Sequence
76	59.8	3.9	168479	2	AC135674 Ratius no
77	59.8	3.9	221787	2	AC119782 Ratius no
78	59.8	3.9	221787	2	AC097926 Ratius no
79	59.8	3.9	246566	6	CQ720332 Sequence
80	59.8	3.8	5208	6	AX555393 Sequence
81	58.8	3.8	2000	6	AX655393 Sequence
82	58.8	3.8	6434	9	AK128741 Homo sapi
83	58.4	3.8	219205	10	AL611985 Mouse DNA
84	58.2	3.8	63	9	D86113 Homo sapien
85	58.2	3.8	4343	10	BC060721 Mus muscu
86	58.2	3.8	116984	2	AC133595 Ratius no
87	58.2	3.8	187718	9	AC104581 Homo sapi
88	57.6	3.7	232984	2	AC111597 Ratius no
89	57.6	3.7	261384	2	AC115415 Ratius no
90	57.6	3.7	278068	2	AC123167 Ratius no
91	56.8	3.7	189081	5	BX511250 Zebrafish
92	56.6	3.7	2803	6	CQ721455 Sequence

93	56.6	3.7	6505	9	AB002296	Homo sapi	AB002296 Homo sapi	166	50.8	3.3	155337	2	AC116408	AC116408 Mus muscu
94	56.4	3.7	5811	3	AF210315	Drosophila	AF210315 Drosophila	167	50.6	3.3	51	11	BY190786	BY190786 egmml6842
95	56.4	3.7	6124	6	CQ583074	Sequence	CQ583074 Sequence	168	50.6	3.3	2000	6	AX655393	AX655393 Sequence
96	56.4	3.7	21747	6	CQ583073	Sequence	CQ583073 Sequence	169	50.6	3.3	201768	5	BX323591	BX323591 Zebrafish
97	56.4	3.7	97538	2	AC017943	Drosophila	AC017943 Drosophila	170	50.4	3.3	155337	2	AC116408	AC116408 Mus muscu
98	56.4	3.7	170939	3	AC007757	Drosophila	AC007757 Drosophila	171	50.4	3.3	163121	5	AL928637	AL928637 Zebrafish
99	56.4	3.7	197071	3	AC007771	Drosophila	AC007771 Drosophila	172	50.4	3.3	170189	2	BX957358	BX957358 Dario rer
100	56.4	3.7	230001	3	AE003731	Drosophila	AE003731 Drosophila	173	50.2	3.2	2151	10	BC025482	BC025482 Mus muscu
101	56.2	3.6	66986	2	AC133881	Homo sapi	AC133881 Homo sapi	174	50.2	3.2	3951	10	S78219	S78219 nuclear pro
102	56.2	3.6	165212	10	AL596125	Mouse DNA	AL596125 Mouse DNA	175	50.2	3.2	4032	10	BC056959	BC056959 Mus muscu
103	56.2	3.6	183887	2	AC084415	Mus muscu	AC084415 Mus muscu	176	50.2	3.2	4053	10	S78221	S78221 nuclear pro
104	55.8	3.6	33957	5	AY598453	Danio rer	AY598453 Danio rer	177	50	3.2	239130	2	AC079420	AC079420 Mus muscu
105	55.8	3.6	5003	6	CQ730634	Sequence	CQ730634 Sequence	178	49.6	3.2	65351	2	AC139773	AC139773 Homo sapi
106	55.4	3.6	56520	9	AC135892	Homo sapi	AC135892 Homo sapi	179	49.6	3.2	170328	8	AP005829	AP005829 Oryza sat
107	55.4	3.6	172571	9	AC006064	Homo sapi	AC006064 Homo sapi	180	49.6	3.2	230387	2	AC051622	AC051622 Mus muscu
108	55.4	3.6	245880	2	AC079387	Homo sapi	AC079387 Homo sapi	181	49.4	3.2	201	11	BY179021	BY179021 egmml0356
109	54.4	3.5	135259	9	HS127820	Human DNA	HS127820 Human DNA	182	49.4	3.2	60032	2	AC129549	AC129549 Homo sapi
110	54.2	3.5	1158	9	HST000230	Homo sapi	AL360186 Homo sapi	183	49.4	3.2	65351	2	AC139773	AC139773 Homo sapi
111	54.2	3.5	2596	9	HST000226	Homo sapi	AL365456 Homo sapi	184	49	3.2	166862	10	AC124457	AC124457 Mus muscu
112	54.2	3.5	3039	9	AF009353	Homo sapi	AP009353 Homo sapi	185	49	3.2	167396	2	CR352259	CR352259 Danio rer
113	54.2	3.5	3799	6	CQ714282	Sequence	CQ714282 Sequence	186	49	3.2	202748	10	MMU307670	MMU307670 Mus muscu
114	54.2	3.5	3807	9	AF119042	Homo sapi	AF119042 Homo sapi	187	49	3.2	300800	1	SC0939112	SC0939112 Sequence
115	54.2	3.5	3909	9	BC028689	Homo sapi	BC028689 Homo sapi	188	48.8	3.2	3079	6	AX880162	AX880162 Sequence
116	54.2	3.5	3997	6	CQ783904	Sequence	CQ783904 Sequence	189	48.8	3.2	3079	6	BD158230	BD158230 Primer fo
117	54.2	3.5	3997	6	BD127893	Primer fo	BD127893 Primer fo	190	48.8	3.2	3079	6	AK023116	AK023116 Homo sapi
118	54.2	3.5	3997	6	AK075306	Homo sapi	AK075306 Homo sapi	191	48.8	3.2	129996	2	AC091664	AC091664 Oryza sat
119	54.2	3.5	136753	2	AC149968	Strongylo	AC149968 Strongylo	192	48.8	3.2	130528	8	AP003202	AP003202 Oryza sat
120	53.8	3.5	300695	2	AC079431	Mus muscu	AC079431 Mus muscu	193	48.8	3.2	136357	8	AP003562	AP003562 Oryza sat
121	53.6	3.5	220674	5	BX548000	Danio rer	BX548000 Danio rer	194	48.8	3.2	178331	2	BX957269	BX957269 Danio rer
122	53	3.4	3116	5	AY598454	Danio rer	AY598454 Danio rer	195	48.6	3.1	1755	6	CQ731955	CQ731955 Sequence
123	52.8	3.4	292	6	CQ079223	Sequence	CQ079223 Sequence	196	48.6	3.1	2251	9	BC078181	BC078181 Homo sapi
124	52.8	3.4	292	6	CQ110776	Sequence	CQ110776 Sequence	197	48.6	3.1	2579	9	AY278022	AY278022 Homo sapi
125	52.8	3.4	292	6	CQ149515	Sequence	CQ149515 Sequence	198	48.6	3.1	2510	9	BC039002	BC039002 Homo sapi
126	52.8	3.4	292	6	CQ184336	Sequence	CQ184336 Sequence	199	48.6	3.1	2630	9	BC038222	BC038222 Homo sapi
127	52.8	3.4	292	6	CQ232797	Sequence	CQ232797 Sequence	200	48.6	3.1	128266	9	AC008623	AC008623 Homo sapi
128	52.8	3.4	292	6	CQ270820	Sequence	CQ270820 Sequence							
129	52.8	3.4	292	6	CQ308111	Sequence	CQ308111 Sequence							
130	52.8	3.4	292	6	CQ344926	Sequence	CQ344926 Sequence							
131	52.8	3.4	476	6	CQ070006	Sequence	CQ070006 Sequence							
132	52.8	3.4	476	6	CQ097609	Sequence	CQ097609 Sequence							
133	52.8	3.4	476	6	CQ136450	Sequence	CQ136450 Sequence							
134	52.8	3.4	476	6	CQ174430	Sequence	CQ174430 Sequence							
135	52.8	3.4	476	6	CQ219756	Sequence	CQ219756 Sequence							
136	52.8	3.4	476	6	CQ258219	Sequence	CQ258219 Sequence							
137	52.8	3.4	476	6	CQ295563	Sequence	CQ295563 Sequence							
138	52.8	3.4	476	6	CQ332152	Sequence	CQ332152 Sequence							
139	52.8	3.4	6025	6	CQ613956	Sequence	CQ613956 Sequence							
140	52.8	3.4	7303	2	AC018273	Drosophila	AC018273 Drosophila							
141	52.8	3.4	7303	6	CQ613955	Sequence	CQ613955 Sequence							
142	52.8	3.4	150884	2	AC009382	Drosophila	AC009382 Drosophila							
143	52.8	3.4	167816	3	AC009381	Drosophila	AC009381 Drosophila							
144	52.8	3.4	250662	3	AE003515	Drosophila	AE003515 Drosophila							
145	52.2	3.4	539	6	CQ497245	Sequence	CQ497245 Sequence							
146	52.2	3.4	539	6	CQ512372	Sequence	CQ512372 Sequence							
147	52.2	3.4	560	11	BY061719	Mus muscu	BY061719 S21P6026							
148	52.2	3.4	67573	6	CQ122760	Mus muscu	AC122760 Mus muscu							
149	52	3.4	1215	6	CQ715851	Sequence	CQ715851 Sequence							
150	52	3.4	1647	9	CR456448	Homo sapi	CR456448 Homo sapi							
151	52	3.4	2651	6	AX833782	Sequence	AX833782 Sequence							
152	52	3.4	2651	9	AX095777	Homo sapi	AX095777 Homo sapi							
153	52	3.4	3498	6	AX747427	Sequence	AX747427 Sequence							
154	52	3.4	3498	9	AK092243	Homo sapi	AK092243 Homo sapi							
155	52	3.4	3584	6	AX747013	Sequence	AX747013 Sequence							
156	52	3.4	3584	9	AK091480	Homo sapi	AK091480 Homo sapi							
157	52	3.4	3694	9	BC012187	Homo sapi	BC012187 Homo sapi							
158	51.8	3.4	3694	10	RNND4P	Homo sapi	X66622 R. norvegicu							
159	51.6	3.3	4871	10	AK129293	Mus muscu	AK129293 Mus muscu							
160	51.6	3.3	5686	10	AY458590	Mus muscu	AY458590 Mus muscu							
161	51.4	3.3	1637	10	BC052348	Mus muscu	BC052348 Mus muscu							
162	51.4	3.3	2273	10	MMU48238	Mus muscu	MMU48238 Mus muscu							
163	51.4	3.3	212974	2	BX927371	Danio rer	BX927371 Danio rer							
164	51.4	3.3	6475	3	AF119716	Drosophila	AF119716 Drosophila							
165	51	3.3	129506	2	AC136100	Rattus no	AC136100 Rattus no							

/map="21q22.3"
/sex="male"
/tissue_type="chymus"
/dev_stage="3-yr-old"
1..1545
/gene="ABECED"
237..1283
/gene="ABECED"
/note="autoimmune regulator-2"
/product="AIRB-2"
/codon_start=1
/protein_id="BAA23989.1"
/db_xref="GI:2696617"
/translation="MNLVYSSGAPGTQDPAKRVFPPIGMARGVCMRPDKGTGGG
RISFSGWAGRLSSGTQRCWSCFEKEVALRVLHPSVCMGVSCLQKNEDEC
AVCRDGGELICDDGCEPAPHLACLSPLEAIPISTGWRCSCLQATQVEQPAEPRP
DEPVEPTLPPLRSGAEVRGPPGEPPLAGMDTLVYKILPAPSPAPLPGDLSALH
PLLCVPEQONILAPGARCVGCGDGTDLRCHCAAFHMRCHPAGTGRPGIGRRC
SGSGVTTPAPVEGVLAPSPARLAPPAKODTASHPALHRDLSELSHTDGLIQW
AIQSMARPAAPFPs"
1531..1536
/gene="ABECED"

polyA_signal

ORIGIN

Query Match 99.8%; Score 1541.8; DB 9; Length 1545;
Beef Local Similarity 99.9%; Pred. No. 4.3e-290;
Matches 1543; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGAGAAAGTAGAGTCTTCTCAGGCTCTTAAGAGCATGGGCTTGTCTCAGGCTGTACCG 60
Db 1 AGAGAAAGTAGAGTCTTCTCAGGCTCTTAAGAGCATGGGCTTGTCTCAGGCTGTACCG 60

QY 61 CTGCTCTCAGCTGGGCGCCGTGGGCGCGGCGCGCCCTGCTATAGCCAGAGGTCAAG 120
Db 61 CTGCTCTCAGCTGGGCGCCGTGGGCGCGGCGCGCCCTGCTATAGCCAGAGGTCAAG 120

QY 121 ATCCATGGGGAATGCCATGCTACTTGTGTGCCAGCATGGTCTTATGGGGTAGAA 180
Db 121 ATCCATGGGGAATGCCATGCTACTTGTGTGCCAGCATGGTCTTATGGGGTAGAA 180

QY 181 GCAGGTCTGGAGAGACCTCTCTGGGCGTGGCCCACTGCTGTGAGAGAGGTTTCATGT 240
Db 181 GCAGGTCTGGAGAGACCTCTCTGGGCGTGGCCCACTGCTGTGAGAGAGGTTTCATGT 240

QY 241 GGTGGTGTACAGTTCCTGGGCGCCCTGGAACCGACAGACCTCTGCAAGAACCGGGTTTCT 300
Db 241 GGTGGTGTACAGTTCCTGGGCGCCCTGGAACCGACAGACCTCTGCAAGAACCGGGTTTCT 300

QY 301 TCCCAATAGGAGTAGGCGCGGGGGGTGTCTGTTCAGACCAATGGAATGGGGAACAGGTG 360
Db 301 TCCCAATAGGAGTAGGCGCGGGGGGTGTCTGTTCAGACCAATGGAATGGGGAACAGGTG 360

QY 361 GTCAAGGGCAGAAATTTCAAGCCCTCTGAGCATGGAGCAGAGCAGAGACTGGGAGTTTCA 420
Db 361 GTCAAGGGCAGAAATTTCAAGCCCTCTGAGCATGGAGCAGAGCAGAGACTGGGAGTTTCA 420

QY 421 GTACCCAGAGATGCTGCTGGGAGAGCTGTTTGGGAAGAGGTGGCTCTCAGAGAGGTGC 480
Db 421 GTACCCAGAGATGCTGCTGGGAGAGCTGTTTGGGAAGAGGTGGCTCTCAGAGAGGTGC 480

QY 481 TGACACCCAGCCAGCTGCTGATGGGGGTCTCTTGGCCAGAGAAATGAGAGAGT 540
Db 481 TGACACCCAGCCAGCTGCTGATGGGGGTCTCTTGGCCAGAGAAATGAGAGAGT 540

QY 541 GTGCGGTGTGTGGGAGCGGCGGAGACTATCTGTGTAGACGAGCTGCTGGGCGCTTCC 600
Db 541 GTGCGGTGTGTGGGAGCGGCGGAGACTATCTGTGTAGACGAGCTGCTGGGCGCTTCC 600

QY 601 ACTGCGCTGCTGTCTCTGCTCGGCTCGGAGAGATCCCAAGTGGAGACTGGAAGTGTCCA 660
Db 601 ACTGCGCTGCTGTCTCTGCTCGGCTCGGAGAGATCCCAAGTGGAGACTGGAAGTGTCCA 660

QY 661 GCTGCTGTGAGCAAGTCTCAGAGAGTGCAGACCCCGGAGAGAGAGCCCGGCCAGG 720

Db 661 GCTGCTGTGAGCAAGTCTCAGAGAGTGCAGACCCCGGAGAGAGAGCCCGGCCAGG 720

QY 721 AGCCACCCGTGAGAGACCCCGCTCCCCCGGAGCTTAAGTGGCGGAGAGAGGTAAAG 780
Db 721 AGCCACCCGTGAGAGACCCCGCTCCCCCGGAGCTTAAGTGGCGGAGAGAGGTAAAG 780

QY 781 GTTCACCTGGGGAACCCCTTACGCGGATGAGACAGACTTTGTCTTCAAGACACTGCGG 840
Db 781 GTTCACCTGGGGAACCCCTTACGCGGATGAGACAGACTTTGTCTTCAAGACACTGCGG 840

QY 841 CTCGCGCTTCTGAGAGCCCGCTGCGAGGTCTGAGACTGCTGCGGCGCTGCACTGCTG 900
Db 841 CTCGCGCTTCTGAGAGCCCGCTGCGAGGTCTGAGACTGCTGCGGCGCTGCACTGCTG 900

QY 901 GTGTGGGTCTGAGAGGTGAGAGAACTTGTCTGTGTGCGGCTTGGCGGAGTGTGGAG 960
Db 901 GTGTGGGTCTGAGAGGTGAGAGAACTTGTCTGTGTGCGGCTTGGCGGAGTGTGGAG 960

QY 961 ATGTATCGGAGCTGCTGCGGTGTACTCACTGCGGCGCTTCTTCACTGGGCTGCACT 1020
Db 961 ATGTATCGGAGCTGCTGCGGTGTACTCACTGCGGCGCTTCTTCACTGGGCTGCACT 1020

QY 1021 TCCAGCGGAGCACTCCCGGCGCGGAGAGCGGCGCTGCGAGATCTGCTGAGAGAG 1080
Db 1021 TCCAGCGGAGCACTCCCGGCGCGGAGAGCGGCGCTGCGAGATCTGCTGAGAGAG 1080

QY 1081 TGACCCAGACCCCTGTGAGAGGGGAGTGTGGCCCGGAGCCCGGCGCTGCGGCG 1140
Db 1081 TGACCCAGACCCCTGTGAGAGGGGAGTGTGGCCCGGAGCCCGGCGCTGCGGCG 1140

QY 1141 CTGCCAAGATGACACTGCTGCACTGAGAGCCGCTCTGCAAGGATGACTGTGAGTCCC 1200
Db 1141 CTGCCAAGATGACACTGCTGCACTGAGAGCCGCTCTGCAAGGATGACTGTGAGTCCC 1200

QY 1201 TTCTGAGGAGAGCACTTGCATGAGTGGATCTGCTGAGAGGAGCACTGAGAGAGTCCC 1260
Db 1201 TTCTGAGGAGAGCACTTGCATGAGTGGATCTGCTGAGAGGAGCACTGAGAGAGTCCC 1260

QY 1261 CGGCGGCGCCCTTCTCTCTCTGAGCCCGAGATGGCGGAGCATGCACTGTATGAGAG 1320
Db 1261 CGGCGGCGCCCTTCTCTCTCTGAGCCCGAGATGGCGGAGCATGCACTGTATGAGAG 1320

QY 1321 TGCTGAGAGAGACACTCTCTCTCTGAGTCTGAGAGCGGCGGCTGAGATGAGAGG 1380
Db 1321 TGCTGAGAGAGACACTCTCTCTCTGAGTCTGAGAGCGGCGGCTGAGATGAGAGG 1380

QY 1381 GACAGCGCACTCTTGTCAAGTGTGCTGAGTGAACAGCTGTGTCTGTGGGACACCA 1440
Db 1381 GACAGCGCACTCTTGTCAAGTGTGCTGAGTGAACAGCTGTGTCTGTGGGACACCA 1440

QY 1441 GCCATCATGTGCTGGAATTAACCTTGCAGTGTGCTGAGTGAACAGCTGTGTCTGTGG 1500
Db 1441 GCCATCATGTGCTGGAATTAACCTTGCAGTGTGCTGAGTGAACAGCTGTGTCTGTGG 1500

QY 1501 GCTCTCTCTGCTGCTGAGCTTCACTTAAATTAATTAATTAATTAATTAATTAATTA 1545
Db 1501 GCTCTCTCTGCTGCTGAGCTTCACTTAAATTAATTAATTAATTAATTAATTAATTA 1545

RESULT 2
CQ731863 1545 bp DNA linear PAT 03-FRB-2004
DEFINITION Sequence 17797 from Patent WO02068579.
ACCESSION CQ731863
VERSION CQ731863.1 GI:42309567
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.

TITLE Kite, such as nucleic acid arrays, comprising a majority of humenexons or transcripts, for detecting expression and other uses thereof
JOURNAL Patent: WO 02068579-A 17797 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES Location/Qualifiers
source 1..1545
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 99.7%; Score 1540.2; DB 6; Length 1545;
Best Local Similarity 99.8%; Pred. No. 8.7e-290;
Matches 1542; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AGAGAAAGTGAAGTCTTCTCAGGCTCTTAAGACATGGGCTTGTCTCAGGCTGTACCCG 60
DB 1 AGAGAAAGTGAAGTCTTCTCAGGCTCTTAAGACATGGGCTTGTCTCAGGCTGTACCCG 60
QY 61 CTGCTCTCAGTGGGCGCGGTGGGTGGGCGCGGCGCGCGCTCTATACCAAGAGTCAAG 120
DB 61 CTGCTCTCAGTGGGCGCGGTGGGTGGGCGCGGCGCGCTCTATACCAAGAGTCAAG 120
QY 121 ATCACTGGGAATGCCATGCTCATCTTTCGTCCCAAGCATGGTTCTTAATGGGGTAGAA 180
DB 121 ATCACTGGGAATGCCATGCTCATCTTTCGTCCCAAGCATGGTTCTTAATGGGGTAGAA 180
QY 181 GCAGGTGCGGAGAGACCTTCCTTGGGCTTGGCCCACTGCTGTGAGAAAGGTTCAAT 240
DB 181 GCAGGTGCGGAGAGACCTTCCTTGGGCTTGGCCCACTGCTGTGAGAAAGGTTCAAT 240
QY 241 GGTGGTGTATACGTTCCGGGGCGCGCTGGAACGAGAGAGCTGTGAAGAAACCGGCTTTCT 300
DB 241 GGTGGTGTATACGTTCCGGGGCGCGCTGGAACGAGAGAGCTGTGAAGAAACCGGCTTTCT 300
QY 301 TCCCAATAGAGATGAGCCCGGCGGGGTGTCTGTTCGAGACAGATGATGGGGAACAGTG 360
DB 301 TCCCAATAGAGATGAGCCCGGCGGGGTGTCTGTTCGAGACAGATGATGGGGAACAGTG 360
QY 361 GTTCAGGGCAAGATTTCAGGCGCTGAGCAATGGAGACAGAGAGCTGGGGAGTTTCA 420
DB 361 GTTCAGGGCAAGATTTCAGGCGCTGAGCAATGGAGACAGAGAGCTGGGGAGTTTCA 420
QY 421 GTTACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAAAGAGAGTGTCTCAAGAGGCTC 480
DB 421 GTTACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAAAGAGAGTGTCTCAAGAGGCTC 480
QY 481 TGCACCCCAAGCCAGTCTGATGAGGCGTCTTTCCTGTGCTCCAGAAATGAGACGAGT 540
DB 481 TGCACCCCAAGCCAGTCTGATGAGGCGTCTTTCCTGTGCTCCAGAAATGAGACGAGT 540
QY 541 GTTCCGTGTGTCTGGAGCGGCGGGAGCTCATCTGTGTGACGCTGCGCTCGGGCTTTC 600
DB 541 GTTCCGTGTGTCTGGAGCGGCGGGAGCTCATCTGTGTGACGCTGCGCTCGGGCTTTC 600
QY 601 ACCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
DB 601 ACCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
QY 661 GGTGCTGTCAGGCAAGTTCAGAGAGTGAAGCCCGGAGAGAGAGCCCGGCGCCAG 720
DB 661 GGTGCTGTCAGGCAAGTTCAGAGAGTGAAGCCCGGAGAGAGAGCCCGGCGCCAG 720
QY 721 AGCCACCCGTGAGACCCCGCTTCCCGGGCTTTCGTGCTGCTGCTGCTGCTGCTGCTG 780
DB 721 AGCCACCCGTGAGACCCCGCTTCCCGGGCTTTCGTGCTGCTGCTGCTGCTGCTGCTG 780
QY 781 GTTCACTGGGGAAACCTTACCGGAGTGAACGACTTGTGTCTACAGACCTTCGCG 840
DB 781 GTTCACTGGGGAAACCTTACCGGAGTGAACGACTTGTGTCTACAGACCTTCGCG 840
QY 841 CTCGCGCTTCTGACGCCCGCTGACAGTCTGAGCTCTGCGCCCTGACCCCTTACTGT 900

DB 841 CTCGCGCTTCTGACGCCCGCTGACAGTCTGAGCTGAATCTCTCGGCTCTGACCCCTTACTGT 900
QY 901 GTTGGGTCTTGAAGGATGACAGAACTTGTCTTGTGTGCGGCTTGTGCGGAG 960
DB 901 GTTGGGTCTTGAAGGATGACAGAACTTGTCTTGTGTGCGGCTTGTGCGGAG 960
QY 961 ATGTATACGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
DB 961 ATGTATACGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
QY 1021 TCCCAAGCGGACACTTCCCGGCGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
DB 1021 TCCCAAGCGGACACTTCCCGGCGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
QY 1081 TGACCCCAAGCCCTGTGAGGGGGTGTGCGGCCCAAGCCCGGCTGCGCTTGGGCT 1140
DB 1081 TGACCCCAAGCCCTGTGAGGGGGTGTGCGGCCCAAGCCCGGCTGCGCTTGGGCT 1140
QY 1141 CTGCAAGGATGACACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 1200
DB 1141 CTGCAAGGATGACACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 1200
QY 1201 TTCTGAGCAGACACTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
DB 1201 TTCTGAGCAGACACTTTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
QY 1261 CGGCGGCGGCTTCCCTCTCTGATCCCAAGATGCGGAGATGCGGATGATGAGAG 1320
DB 1261 CGGCGGCGGCTTCCCTCTCTGATCCCAAGATGCGGAGATGCGGATGATGAGAG 1320
QY 1321 TGTGTAGAGAGACACTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1380
DB 1321 TGTGTAGAGAGACACTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1380
QY 1381 GACAGGCGCACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
DB 1381 GACAGGCGCACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
QY 1441 GCCATCATGTGCTGGAATTTAAACCTGCGCCCACTTCTCTGATGATGATGATGATG 1500
DB 1441 GCCATCATGTGCTGGAATTTAAACCTGCGCCCACTTCTCTGATGATGATGATGATG 1500
QY 1501 GCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1545
DB 1501 GCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1545
RESULT 3
AB006685 1463 bp mRNA linear PRI 13-DEC-1997
LOCUS AB006685
DEFINITION Homo sapiens APECD mRNA for AIRE-3, complete cds.
ACCESSION AB006685.1 GI:2696622
VERSION AB006685.1
KEYWORDS APECD; AIRE-3.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Nagamine,K., Peterson,P., Scott,H.S., Kudoh,J., Minoshima,S.,
Heino,M., Krohn,K.J.E., Laijot,M.D., Mullis,P.E.,
Antonarakis,S.E., Kawasaki,K., Asakawa,S., Ito,F. and Shimizu,N.
TITLE Positional cloning of the APECD gene
JOURNAL Nat. Genet. 17 (4), 393-398 (1997)
MEDLINE 9398839
PUBMED 78061086
REFERENCE
AUTHORS Shimizu,N.
TITLE Direct Submision
JOURNAL Submitted (16-AUG-1997) Nobuyoshi Shimizu, Keio University School
of Medicine, Department of Molecular Biology, 35 Shinanomachi,
Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@mb.med.keio.ac.jp).

FEATURES
Tel:03-3351-2370, Fax:03-3351-2370)

Location/Qualifiers

1..1463

source
/organism="Homo sapiens"

/mol_type="mRNA"

/isolate="Caucasian"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/sex="male"

/tissue_type="thymus"

/dev_stage="3-yr-old"

1..1463

/gene="APECD"

237..1001

/gene="APECD"

/note="autoimmune regulator-3"

/codon_start=1

/product="AIRE-3"

/protein_id="BA023993.1"

/db_xref="GI:2696623"

/translation="MNLVYSSGAPGTQAPNRVFPPIGMAPGVCMRPDGMGTGOG
RISPGSMAGORLSSGTORCCMSGFGKEVALRVLPSPVCMGVSCICQKNEDEC
AVRODGEILICDGPAPHLAICTSPLEIPSTWRGSSCIATQVQVOPRAEERBP
QEPVETPLPBCIRASGEPRCCGATPRCTPTCVWLVKSRMLVRAVAGCAEMVRT
CCGVLTAPLPSTGATSTQPAPEPGRACADPAQET"

1449..1454

/gene="APECD"

polya_signal

ORIGIN

Query Match 88.5%; Score 1367.8; DB 9; Length 1463;

Best Local Similarity 94.6%; Pred. No. 3.3e-256;

Matches 1461; Conservative 0; Mismatches 2; Indels 82; Gaps 1;

1 AGAGAAAGTGAAGTCTTCTCAGGCTCTTAAGACATGCGCTTGTCTCAGGCTGTACCCG 60
1 AAGAGAAAGTGAAGTCTTCTCAGGCTCTTAAGACATGCGCTTGTCTCAGGCTGTACCCG 60
61 GTGCTCTCAGCTGGGCGCCGCTGGGCGCGGCGCCCTGCTAATAGCCAGAGAGTCAAG 120
61 CTGCTCTCAGCTGGGCGCCGCTGGGCGCGGCGCCCTGCTAATAGCCAGAGAGTCAAG 120
121 ATCCACTGGGAATGCCATGCTCATCTTTGTCTCCCGACATGCTTTCTTAATGGGGTAA 180
121 ATCCACTGGGAATGCCATGCTCATCTTTGTCTCCCGACATGCTTTCTTAATGGGGTAA 180
181 GCAGGTCGGGAGAGACTCTCTGAGGCTGGGCGCCCACTGCTGTGAGAAAGGTTCAAT 240
181 GCAGGTCGGGAGAGACTCTCTGAGGCTGGGCGCCCACTGCTGTGAGAAAGGTTCAAT 240
241 GGTGTGTGTACAGTTCCGGGCGCCCTGGAACGACAGACCTGCAAGAAACCGGCTTTCT 300
241 GGTGTGTGTACAGTTCCGGGCGCCCTGGAACGACAGACCTGCAAGAAACCGGCTTTCT 300
301 TCCCAATAGGAATGGGCGCGCGGGGCTCTGTTCCAGACCAAGATGATGGGAAACAG 360
301 TCCCAATAGGAATGGGCGCGCGGGGCTCTGTTCCAGACCAAGATGATGGGAAACAG 360
361 GTCAAGGCGAGAAATTTTCAAGCCCTGACAGACATGGGCGAGAGACTGGGAGTTTCA 420
361 GTCAAGGCGAGAAATTTTCAAGCCCTGACAGACATGGGCGAGAGACTGGGAGTTTCA 420
421 GTACCCAGAGATGCTCTGGGCGAGCTGTTTGGAAAGAGTGCTCTCAGAGAGGTGC 480
421 GTACCCAGAGATGCTCTGGGCGAGCTGTTTGGAAAGAGTGCTCTCAGAGAGGTGC 480
481 TGGACCCCGCCAGCTGATGAGGAGCTCTTGGCTCTGAGCCAGAGAAATGAGAGAGT 540
481 TGGACCCCGCCAGCTGATGAGGAGCTCTTGGCTCTGAGCCAGAGAAATGAGAGAGT 540
541 GTGCGCTGTGTGCGGAGCGGCGGAGACTCATCTGTGACGGCTCCCTCGGCGCTTC 600
541 GTGCGCTGTGTGCGGAGCGGCGGAGACTCATCTGTGACGGCTCCCTCGGCGCTTC 600

601 ACCTGGCTGTGCTGTCTCCCTCCGCTCCGAGAGATCCCAAGTGGAGCTTGAGGTCTCCA 660
601 ACCTGGCTGTGCTGTCTCCCTCCGCTCCGAGAGATCCCAAGTGGAGCTTGAGGTCTCCA 660
661 GCTGCTGAGGAGCAAGTCCAGAGAGTCCAGAGAGTCCGAGAGAGAGAGAGAGAGAG 720
661 GCTGCTGAGGAGCAAGTCCAGAGAGTCCAGAGAGTCCGAGAGAGAGAGAGAGAGAG 720
721 AGCCACCCGTGAG 780
721 AGCCACCCGTGAG 780
781 GTCCACCTGGGAG 840
781 GTCCACCTGGGAG 840
841 CTCGCGCTTCTGAGAGCGCGCTGAGAGTGTGAGAGTCTGAGAGAGAGAGAGAGAG 900
841 CTCGCGCTTCTGAGAGCGCGCTGAGAGTGTGAGAGTCTGAGAGAGAGAGAGAGAG 900
901 GTGCGCTGCTGAG 960
901 GTGCGCTGCTGAG 960
961 ATGTGACGAG 1020
961 ATGTGACGAG 1020
1021 TCCAGCGCGAG 1080
1021 TCCAGCGCGAG 1080
939 TCCAGCGCGAG 998
939 TCCAGCGCGAG 998
1081 TGACCCCAAG 1140
1081 TGACCCCAAG 1140
999 TGACCCCAAG 1058
999 TGACCCCAAG 1058
1141 CTGCGAG 1200
1141 CTGCGAG 1200
1059 CTGCGAG 1118
1059 CTGCGAG 1118
1201 TTCTGAG 1260
1201 TTCTGAG 1260
1119 TTCTGAG 1178
1119 TTCTGAG 1178
1261 CGGCG 1320
1261 CGGCG 1320
1179 CGGCG 1238
1179 CGGCG 1238
1321 TGTGAG 1380
1321 TGTGAG 1380
1239 TGTGAG 1298
1239 TGTGAG 1298
1381 GACAGCGAG 1440
1381 GACAGCGAG 1440
1299 GACAGCGAG 1358
1299 GACAGCGAG 1358
1441 GCCATCATGTGCTGAGAAATTAACCTGCGCCCACTTCTCTAATCTGAGAGTCCCGG 1500
1441 GCCATCATGTGCTGAGAAATTAACCTGCGCCCACTTCTCTAATCTGAGAGTCCCGG 1500
1359 GCCATCATGTGCTGAGAAATTAACCTGCGCCCACTTCTCTAATCTGAGAGTCCCGG 1418
1359 GCCATCATGTGCTGAGAAATTAACCTGCGCCCACTTCTCTAATCTGAGAGTCCCGG 1418
1501 GCTCTCTCTGCTGCTGAGCTTAAATTAATTAATTAATTAATTAATTAATTAATTA 1545
1501 GCTCTCTCTGCTGCTGAGCTTAAATTAATTAATTAATTAATTAATTAATTAATTA 1545
1419 GCTCTCTCTGCTGCTGAGCTTAAATTAATTAATTAATTAATTAATTAATTAATTA 1463
1419 GCTCTCTCTGCTGCTGAGCTTAAATTAATTAATTAATTAATTAATTAATTAATTA 1463

RESULT 4

AB006682

LOCUS Homo sapiens APECD mRNA for AIRE-1, complete cds. 2027 bp mRNA linear PRI 13-DEC-1997

DEFINITION AB006682

AB006682.1 GI:2696614

AB006682.1 GI:2696614

AB006682.1 GI:2696614

AB006682.1 GI:2696614

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Nagamine,K., Peterson,P., Scott,H.S., Kudoh,J., Minoshima,S., Heino,M., Krohn,K.J.E., Lalioti,M.D., Mullis,P.E., Antonarakis,S.E., Kawasaki,K., Asakawa,S., Ito,F. and Shimizu,N.
TITLE Positional cloning of the APECED gene
JOURNAL Nat. Genet. 17 (4), 393-398 (1997)
MEDLINE 98061086
PUBMED 9398839
REFERENCE 2 (bases 1 to 2027)
AUTHORS Shimizu,N.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-1997) Nobuyoshi Shimizu, Keio University School of Medicine, Department of Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@mb.med.keio.ac.jp, Tel:03-3351-2370, Fax:03-3351-2370)
Location/Qualifiers
FEATURES
source 1..2027
/organism="Homo sapiens"
/mol_type="mRNA"
/isolate="Caucasian"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/sex="male"
/tissue_type="thymus"
/dev_stage="3-yr-old"
1..2027
/gene="APECED"
128..1765
/note="autoimmune regulator-1"
/product="AIRE-1"
/protein_id="BAA23988.1"
/db_xref="GI:2696615"
/translation="MATDAALRLRLALHRTETAVAVDSAPFLHIALADHDVPEKQ
ETLHKKEGCEPQAFHALISWLTQDSTAILDPRVLPKQINLERYGLQPLIDSPK
DVLDSQRRGRKRPVPAKLVPPRLPTRRKASEARAAAPALTRYGTASPSQAKA
KPKPKRESSAEOQLPLNGIQTMSASVORAVMSGDVPAAGAVEGLIQVFESG
GSKKCIQVGEFTYTPSKFEDSGSKNKASSSPKPLVAKAGQAPGSGEIRLQO
GSVPAPLPSDPLQHKMEDECAVCRDGBELICCGCPRAFLIACLSPIRISGT
WRCSSCIQATVQVOPRAEPPVPTPLPGIRSGEVRGPRGRLMGMDTL
VYKHLPAPPSAPLPGIDSSALHPLICVPPBGGQNALPARGCVCDGTDVLRCTCA
AALHMRCHPAGTSRPTGLRCKSCSGDVPAPVBSVLAPSPARLAPGPAKDDTASHE
PALHRDLSLSLSEHTFDGILQWALQSMARPAAPFS"
2013..2018
/gene="APECED"

polYA_signal
ORIGIN
Query Match 66.3%; Score 1023.6; DB 9; Length 2027;
Best Local Similarity 99.6%; Pred. No. 3.3e-189;
Matches 1026; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 516 CTGTGCAGAGAATGAGAGCAGAGTGCCTGTGTGCGGAGCGCGGAGGCTCATCTGC 575
DB 998 CTCACACAGAGATAGAGAGAGTGTGCGGTGTGCGGAGCGCGGAGGCTCATCTGC 1057
QY 576 TGTGACGGCTGCGCTGCGGCTTTCACCTGAGCTGCTGCTCCCTCGCTCGGAGAGT 635
DB 1058 TGTGACGGCTGCGCTGCGGCTTTCACCTGAGCTGCTGCTCCCTCGCTCGGAGAGT 1117
QY 636 CCCAGTGGAGCTGAGAGTGTCTCAAGTGTCTGACAGGCAACAGTCCAGAGGTGACGCC 695
DB 1118 CCCAGTGGAGCTGAGAGTGTCTCAAGTGTCTGACAGGCAACAGTCCAGAGGTGACGCC 1177
QY 696 CGGCGAGAGAGCGCGCGCGCGCGAGAGCAACCGGTGAGAGCGCGCTCCCGCGGAGCTT 755
DB 1178 CGGCGAGAGAGCGCGCGCGCGCGAGAGCAACCGGTGAGAGCGCGCTCCCGCGGAGCTT 1237

QY 756 AGGTGCGCGGAGAGAGAGGTAAAGGTCCACCTGCGGAAACCCCTAGCCCGGACGTGACAG 815
DB 1238 AGGTGCGCGGAGAGAGAGGTAAAGGTCCACCTGCGGAAACCCCTAGCCCGGACGTGACAG 1297
QY 816 ACTTGTGTCTACAGACACTGCGGCTCGCTTCTGACGCCCGCGCTGCAAGTGTGAC 875
DB 1298 ACTTGTGTCTACAGACACTGCGGCTCGCTTCTGACGCCCGCGCTGCAAGTGTGAC 1357
QY 876 TCCTGCGCGCTGACACCCCTACTGTGTGTGGTTCCTGAGAGGTGACAGAACCTGGCTCT 935
DB 1358 TCCTGCGCGCTGACACCCCTACTGTGTGTGGTTCCTGAGAGGTGACAGAACCTGGCTCT 1417
QY 936 GGTGCGCTGTGCGGAGTGTGCGAGATGTACAGACGTCTCGGTGTACTCATCTGCGCC 995
DB 1418 GGTGCGCTGTGCGGAGTGTGCGAGATGTACAGACGTCTCGGTGTACTCATCTGCGCC 1477
QY 996 GCTGCTTTCACACTGCGGCTGCGCACTTCCAGCGCGCACTCCCGCGCGGAGCGGCTG 1055
DB 1478 GCTGCTTTCACACTGCGGCTGCGCACTTCCAGCGCGCACTCCCGCGCGGAGCGGCTG 1537
QY 1056 CGCTGAGATCTCTGCTCAGAGACGTGACCCGAGCCCTGTGAGAGGGGTGCTGCGCC 1115
DB 1538 CGCTGAGATCTCTGCTCAGAGACGTGACCCGAGCCCTGTGAGAGGGGTGCTGCGCC 1597
QY 1116 AGCCCGCGCGCTGCGCTGCGCTGCGCAAGATGACACTGCGCAAGTCCAGAGCCGCT 1175
DB 1598 AGCCCGCGCGCTGCGCTGCGCTGCGCAAGATGACACTGCGCAAGTCCAGAGCCGCT 1657
QY 1176 CTGACAGAGATGACCTGAGATCTCTTCTGACGAGACACACTTGTGATGGCATCTGCG 1235
DB 1658 CTGACAGAGATGACCTGAGATCTCTTCTGACGAGACACACTTGTGATGGCATCTGCG 1717
QY 1236 TGGGCGCATCAGAGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1295
DB 1718 TGGGCGCATCAGAGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1777
QY 1296 GGGACATGACGCTGTGATGAGAGTGTGAGAGAGACACTTCTTCTCAGTCTGAGAA 1355
DB 1778 GGGACATGACGCTGTGATGAGAGTGTGAGAGAGACACTTCTTCTCAGTCTGAGAA 1837
QY 1356 GCGCGCGCGCTGAGATCAAGAGGGGACAGCGCACCTTGTCAATGCTCGGCTTAA 1415
DB 1838 GCGCGCGCGCTGAGATCAAGAGGGGACAGCGCACCTTGTCAATGCTCGGCTTAA 1897
QY 1416 CAGCTGTGTGTTTCGAGGACACAGCATCATGTGCTGGAATTAACCTGCGCCAC 1475
DB 1898 CAGCTGTGTGTTTCGAGGACACAGCATCATGTGCTGGAATTAACCTGCGCCAC 1957
QY 1476 TTCTTACTCTGGAAGTCCCGGAGCGCTCTCTTGTGCTGAGCTTAAATATTA 1535
DB 1958 TTCTTACTCTGGAAGTCCCGGAGCGCTCTCTTGTGCTGAGCTTAAATATTA 2017
QY 1536 AAATTAGCTG 1545
DB 2018 AAATTAGCTG 2027

RESULT 5
A97284
LOCUS A97284 2245 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 1 from Patent WO918197.
ACCESSION A97284
VERSION A97284.1 GI:6780668
KEYWORDS
SOURCE unidentifed
ORGANISM unidentifed
REFERENCE 1 (bases 1 to 2245)
AUTHORS Yasuo,M. and Lehmach,H.
TITLE NUCLEIC ACID MOLECULE ENCODING A (POLY)PEPTIDE CO-SEGREGATING IN MUTATED FORM WITH AUTOIMMUNE POLYENDOCRINOPATHY CANDIDIASIS
JOURNAL ECTODERMAL DYSTROPHY (APECED)
Patent: WO 918197-A, 1 15-Apr-1999;

[illegible]

Query Match	Best Local Similarity	Matches	Score	DB	Length	20000;
29.9%;	97.7%;	0;	462.4;	9;	20000;	
Best Local Similarity	97.7%;	Pred. No. 4.3e-80;				
Matches	469;	Conservative	0;	Mismatches	11;	Indels
						Gaps
						0;
183	AGGTCGGAGAGACCTCCCTGGGCGCTGGCCCACTGCCTGTGAGGAAGGTTTCATGTGG	242				
Db	9453	AGGTCGGAGAGACCTCCCTGGGCGCTGGCCCACTGCCTGTGAGGAAGGTTTCATGTGG	9512			
Qy	243	TTGGTGTACAGTTCCGGGGGCCCCCTGGAAACGACAGAGCTTCGACAAACCGGGTTTCTTC	302			
Db	9513	TTGGTGTACAGTTCCGGGGGCCCCCTGGAAACGACAGAGCTTCGACAAACCGGGTTTCTTC	9572			
Qy	303	CCAAATAGGAGATGGCCCGGGGGGGTCTGTTTCAAGACAGATTGATGGGGAACAGGTGGT	362			
Db	9573	CCAAATAGGAGATGGCCCGGGGGGGTCTGTTTCAAGACAGATTGATGGGGAACAGGTGGT	9632			
Qy	363	CAGGCGAATTTCAAGCCCTTGGACAGATGGAGACAGGGCAGAGACTTGGGAGTTTCAGT	422			

Db		9633	CAGGGCAGAATTTCAGGCGCCCTGCAGCATGTGGACACAGGCAAGACTGCGGGAGTTTCAAGT	96932
Oy		423	ACCACGAGATGCTGCTG3GGGAGCTGTTTTGGAGAGAGTGCGCTCTCAGAGGGTGTCTG	482
Db		9693	ACCACGAGATGCTGCTG3GGGAGCTGTTTTGGAGAGAGTGCGCTCTCAGAGGGTGTCTG	97522
Oy		483	CACCCGAGGCCAGCTGCATGAGGAGCGTCTTGGCTGCGCCAGAAATGAGACAGATGT	542
Db		9753	CACCCGAGGCCAGCTGCATGAGGAGCGTCTTGGCTGCGCCAGAAATGAGACAGATGT	98122
Oy		543	GCCGTGTGTGTGGAGCGGCGGGAGCTCATCTGTGTGACGAGCTGCGCTCGGAGCTTTCAC	602
Db		9813	GCCGTGTGTGTGGAGCGGCGGGAGCTCATCTGTGTGACGAGCTGCGCTCGGAGCTTTCAC	98722
Oy		603	CTGGCCGTCCTTCCTCCCTCCGCTCCGGGAGATATCCCAGTGGGACGTGAAGTGTCTCAGC	662
Db		9873	CTGGCCGTCCTTCCTCCCTCCGCTCCGGGAGATATCCCAGTGGGACGTGTGACCTTCAGC	99322
RESULT 9				
HSAJ9610				
LOCUS	HSJ9610	36284 bp	DNA	linear PRI 01-SEP-2000
DEFINITION	Homo sapiens AIRE gene.			
ACCESSION	AJ009610			
VERSION	AJ009610.1 GI:3392939			
KEYWORDS	Aire gene.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1			
REFERENCE	Mattlonen,J., Bjirves,P., Perhentupa,J., Horelli-Kuitunen,N., Palot,A., Peltanen,L., Lee,Y.S., Francis,F., Hennig,S., Thiel,C., Lehrach,H. and Yaspo,M.L. An autoimmune disease, APCED, caused by mutations in a novel gene featuring two PRD-type zinc finger domains Nat. Genet. 17, 399-403 (1997) 2 (bases 1 to 36284) Lee,Y.S., Francis,F., Hennig,S., Thiel,C., Reinhard,R., Lehrach,H. and Yaspo,M.L. Direct Submission Submitted (17-JUL-1998) Streffen Hennig, MPIWG, Abt. Lehrach, Max Planck Institut fuer Molekulare Genetik, Ihnestraße 73, Berlin, 14195, Germany Related cDNA sequence: Z97990. Location/Qualifiers 1..36284 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="21" /map="q22.3" /clone_id="Lawrence Livermore Cosmid 22G11" 199..323 /note="JM3"			
TITLE	JOURNAL	repeat_region	/note="AluY"	complement(1779..10930)
REFERENCE	MEDLINE	repeat_region	/note="AluY"	complement(1101..1198)
AUTHORS		repeat_region	/note="AluSc"	complement(11245..1533)
		repeat_region	/note="AluSg"	complement(11537..1678)
		repeat_region	/note="AluY"	complement(1668..1823)
		repeat_region	/note="AluSc"	complement(1824..2106)
		repeat_region	/note="AluSg"	complement(2112..2196)
		repeat_region	/note="AluSg/X"	complement(2208..2234)
		repeat_region	/note="(GAAA)n"	complement(2235..2532)

repeat_region	/note="AluSx"	complement(3109. .3382)	gene	22529. .34854
repeat_region	/note="AluY"	complement(5332. .5634)	exon	/gene="AIRE"
repeat_region	/note="AluSg"	complement(5722. .5820)		22529. .22779
repeat_region	/note="(GAAA)n"	6217. .6515		/gene="AIRE"
repeat_region	/note="AluSg"	6531. .6761	CDS	/number=1
repeat_region	/note="AluO"	complement(6805. .7106)		join(22648. .22779,23198. .23372,23619. .23774,24158. .24232,
repeat_region	/note="AluSp"	complement(7107. .7409)		24986. .25099,26298. .26443,26629. .26709,27336. .27851,
repeat_region	/note="AluSp"	7445. .7729		28943. .29042,29633. .29815,30428. .30459,31040. .31142,
repeat_region	/note="AluSg"	complement(7757. .7798)		33022. .33084,34291. .34362)
repeat_region	/note="(GAAA)n"	complement(7799. .8101)		/gene="AIRE"
repeat_region	/note="AluSx"	10206. .10439		/codon start=1
repeat_region	/note="Al"	10443. .10742		/protein_id="CAA08759.1"
repeat_region	/note="AluSx"	complement(10918. .11215)	exon	/db_xref="GI:3392940"
repeat_region	/note="AluSx"	complement(11501. .11878)		/db_xref="GDB:O43918"
repeat_region	/note="MSTB"	complement(11993. .12079)	exon	/translation="MATDAALRLRLHRTETIAVVDNAPFLHLADHDVVEBDFQ
repeat_region	/note="MSTB"	12193. .12880	exon	ETLHLKEKGCPOAFHALSLMLTODSTALIDFMRVLFQDYNLERGRLOPLIDSPK
repeat_region	/note="Al"	complement(12887. .13178)	exon	DVDLSOPRRGRKRPVAPKALVPPRLPTFRKASEBARAPALPGRGASGOLKA
repeat_region	/note="AluSc"	13289. .13582		KPKKDESAEQOQLPLNGIOTMSASVORAVAMSGVPGAGAVEGILIOVPESG
repeat_region	/note="AluSx"	13585. .13738	exon	GSKKCIQVGEFTTPTSKTEDSGGKKARSSGPKPLVRAKAQANPQGGARLQO
repeat_region	/note="LI"	complement(13751. .14060)		GSPAPLALPSPDQLQKNEDECAVCRDGLILCCDGCRAFLACLSPLELIPBGT
repeat_region	/note="LI"	complement(14061. .14357)	exon	WRCSSCLQATVOEVQPRAEPRPOEPPVETPLPGLRSAGEVRGPPGRLGMDTL
repeat_region	/note="AluY"	complement(14360. .14720)		VYKHLPAPESAAPLPGLDSSALHPILCVGEGQONLAPARGVMTGLRCRSCGVT
repeat_region	/note="LI"	complement(14722. .15018)	exon	PAVVGVLAPSPARLAPGPAKDTSHEPALHRDLESLSHTTGGTLQWMIQSWAR
repeat_region	/note="AluY"	complement(15019. .15394)		PAAPPS"
repeat_region	/note="LI"	complement(15398. .15700)	exon	23198. .23372
repeat_region	/note="AluSg"	complement(15704. .15864)		/gene="AIRE"
repeat_region	/note="LI"	15866. .16182	exon	/number=2
repeat_region	/note="AluSg/x"	16726. .16931	repeat_region	23619. .23774
repeat_region	/note="MLTIA1"	complement(16933. .17084)	repeat_region	/gene="AIRE"
repeat_region	complement(17090. .17375)		exon	/number=7
repeat_region	/note="AluSx"	complement(17379. .17530)		24158. .24232
repeat_region	/note="MLTIA1"	17663. .17959	exon	/gene="AIRE"
repeat_region	/note="AluY"	18738. .18920	exon	/number=4
repeat_region	/note="MIR"	complement(19498. .20418)		24986. .25099
repeat_region	/note="LTR5"	complement(21877. .21968)	repeat_region	/gene="AIRE"
repeat_region	/note="MIR"	complement(22524. .22642)	repeat_region	/number=10
repeat_region	/note="GC_rich"		repeat_region	/gene="AIRE"
			repeat_region	/number=11
			repeat_region	complement(30872. .30983)
			exon	/note="POLY G"
			exon	31040. .31152
			exon	/gene="AIRE"
			exon	/number=12
			exon	33022. .33084
			exon	/gene="AIRE"
			exon	/number=13
			exon	34291. .34854
			exon	/gene="AIRE"
			exon	/number=14
			exon	34600. .34787
			exon	/note="AluSg"
			exon	complement(34793. .35417)


```
/note="L1MB3"
repeat_region complement(35423..35747)

Query Match      29.9%; Score 462.4; DB 9; Length 36284;
Best Local Similarity 97.7%; Pred. No. 3.9e-80;
Matches 469; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 183 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGCTGAGAGAGGTTCAATGTG 242
DB 27394 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGCTGAGAGAGGTTCAATGTG 27453
QY 243 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 302
DB 27454 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 27513
QY 303 CCAATAGGAGATGGCCCCCGGGGGGTGTCTGTTCGAGACCAAGATGGGGAAACAGGTGT 362
DB 27514 CCAATAGGAGATGGCCCCCGGGGGGTGTCTGTTCGAGACCAAGATGGGGAAACAGGTGT 27573
QY 363 CAGGGCAGAAATTTCAAGCCCTGGCAGCATGGAGCAGGGCAGAGACTGGGGAGTTCAAGT 422
DB 27574 CAGGGCAGAAATTTCAAGCCCTGGCAGCATGGAGCAGGGCAGAGACTGGGGAGTTCAAGT 27633
QY 423 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAGAGAGTGGCTCTCAGAGGGTCTG 482
DB 27634 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAGAGAGTGGCTCTCAGAGGGTCTG 27693
QY 483 CACCCAGCCCACTGTGCAATGGGCGTCTCTTGCTCTGCAAGAAATGAGACGAGTGT 542
DB 27694 CACCCAGCCCACTGTGCAATGGGCGTCTCTTGCTCTGCAAGAAATGAGACGAGTGT 27753
QY 543 GCCGTGTGTCGGGACGCGCGGGAGCTCATCTGCTGTGAGCGGCTCGGCGCTTCCAC 602
DB 27754 GCCGTGTGTCGGGACGCGCGGGAGCTCATCTGCTGTGAGCGGCTCGGCGCTTCCAC 27813
QY 603 CTGGCTGTGCTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGGACCTGAGAGTCTCCAC 662
DB 27814 CTGGCTGTGCTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGGACCTGAGAGTCTCCAC 27873

RESULT 10
AP001060 42133 bp DNA linear PRI 14-APR-2000
LOCUS Homo sapiens genomic DNA, chromosome 21, clone:D4G11, MX1-D21S171
DEFINITION region, complete sequence.
ACCESSION AP001060
VERSION AP001060.1 GI:6693610
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 42133)
Shimizu,N., Kudoh,J. and Shibuya,K.
Homo sapiens genomic DNA, chromosome 21, clone D4G11, MX1-D21S171
TITLE
REFERENCE
AUTHORS
JOURNAL
PUBLISHED Only in Database (2000)
2 (bases 1 to 42133)
Shimizu,N., Kudoh,J. and Shibuya,K.
Direct Submision
Submitted (12-JAN-2000) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail:nshimizu@db.med.keio.ac.jp.
Tel:81-3-3351-2370, Fax:81-3-3351-2370)
FEATURES
source
1..42133
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/clone="D4G11"
/cell_line="GM130B"
/cell_type="B-lymphoblastoid"
```

```
/clone_lib="KU21-D library"
ORIGIN

Query Match      29.9%; Score 462.4; DB 9; Length 42133;
Best Local Similarity 97.7%; Pred. No. 3.8e-80;
Matches 469; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 183 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGCTGAGAGAGGTTCAATGTG 242
DB 30870 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGCTGAGAGAGGTTCAATGTG 30929
QY 243 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 302
DB 30930 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 30989
QY 303 CCAATAGGAGATGGCCCCCGGGGGGTGTCTGTTCGAGACCAAGATGGGGAAACAGGTGT 362
DB 30990 CCAATAGGAGATGGCCCCCGGGGGGTGTCTGTTCGAGACCAAGATGGGGAAACAGGTGT 31049
QY 363 CAGGGCAGAAATTTCAAGCCCTGGCAGCATGGAGCAGGGCAGAGACTGGGGAGTTCAAGT 422
DB 31050 CAGGGCAGAAATTTCAAGCCCTGGCAGCATGGAGCAGGGCAGAGACTGGGGAGTTCAAGT 31109
QY 423 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAGAGAGTGGCTCTCAGAGGGTCTG 482
DB 31110 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAGAGAGTGGCTCTCAGAGGGTCTG 31169
QY 483 CACCCAGCCCACTGTGCAATGGGCGTCTTTCCTCTGTGCAAGAAATGAGACGAGTGT 542
DB 31170 CACCCAGCCCACTGTGCAATGGGCGTCTTTCCTCTGTGCAAGAAATGAGACGAGTGT 31229
QY 543 GCCGTGTGTCGGGACGCGCGGGAGCTCATCTGCTGTGAGCGGCTCGGCGCTTCCAC 602
DB 31230 GCCGTGTGTCGGGACGCGCGGGAGCTCATCTGCTGTGAGCGGCTCGGCGCTTCCAC 31289
QY 603 CTGGCTGTGCTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGGACCTGAGAGTCTCCAC 662
DB 31290 CTGGCTGTGCTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGGACCTGAGAGTCTCCAC 31349

RESULT 11
AC003656_5
WPCOMMENT
Sequence split into 8 fragments LOCUS AC003656 Accession AC003656
Fragment Name Begin End
AC003656_1 100001 210000
AC003656_2 200001 310000
AC003656_3 300001 410000
AC003656_4 400001 510000
AC003656_5 500001 610000
AC003656_6 600001 710000
AC003656_7 700001 753000
Continuation (6 of 8) of AC003656 from base 500001 (AC003656 Homo sapiens clone PI C124)

Query Match      29.9%; Score 462.4; DB 2; Length 110000;
Best Local Similarity 97.7%; Pred. No. 3.3e-80;
Matches 469; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 183 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGCTGAGAGAGGTTCAATGTG 242
DB 52370 AGGTGGGAGAGACCTCCCTGGGCTGGGCCCACTGCTGAGAGAGGTTCAATGTG 52429
QY 243 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 302
DB 52430 TTGGTGTACAGTTCCGGGGCCCTGGAGACGACAGCTTCAGAAACCGGGTTTCTTC 52489
QY 303 CCAATAGGAGATGGCCCCCGGGGGGTGTCTGTTCGAGACCAAGATGGGGAAACAGGTGT 362
DB 52490 CCAATAGGAGATGGCCCCCGGGGGGTGTCTGTTCGAGACCAAGATGGGGAAACAGGTGT 52549
QY 363 CAGGGCAGAAATTTCAAGCCCTGGCAGCATGGAGCAGGGCAGAGACTGGGGAGTTCAAGT 422
```

Db	52550	CAGGGCAGAATTTTCAGGCGCTTGCAGCATGTGGACAGGGCAGAGACTGGGGAGTTCCAGGT	52603
Oy	423	ACCACGAGATGCTCTCTGGGGAGCTGTTTTGGAGAGAGTGCTCTCAGAGGGTGTCTG	482
Db	52610	ACCCAGAGATGCTGCTCTGGGGAGAGCTGTTTTGGAGAGAGTGCTCTCAGAGGGTGTCTG	52663
Oy	483	CACCCAGGCCAGCTGTCATGGGGGTCTCTTGGCTGTGCCAAGAAATGAGACGAGTGT	542
Db	52670	CACCCAGGCCAGCTGTCATGGGGGTCTCTTGGCTGTGCCAAGAAATGAGACGAGTGT	52723
Oy	543	GCCGTGTGTGGGACGCGCGGGAGCTCATCTGCTGTGACCGGCTGCCCCCTTCGAC	602
Db	52730	GCCGTGTGTGGGACGCGCGGGAGCTCATCTGCTGTGACCGGCTGCCCCCTTCGAC	52789
Oy	603	CTGGCCTGTGCTGTCTCCCTCCGCTTCGGGAGATCCCAGTGGGACCTGGAGGTGCTCCAC	662
Db	52790	CTGGCCTGTGCTGTCTCCCTCCGCTTCGGGAGATCCCAGTGGAGCTTCGACCTTCGACG	52849
RESULT 12			
AP001754			
LOCUS	AP001754	340000 bp	DNA linear PRI 06-JUN-2003
DEFINITION	Homo sapiens genomic DNA, chromosome 21q, section 98/105.		
ACCESSION	AP001754 AL163299 BA000005		
VERSION	AP001754.1 GI:7768775		
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Petersson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A., Zimmerman,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K., Asakawa,S., Shintani,A., Saeki,K., Nagamine,K., Mitsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Norisale,G., Hornischler,K., Barand,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Henrich,S., Riesselmann,L., Dargatzis,E., Wehmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L. The DNA sequence of human chromosome 21 Nature 405 (6784), 311-319 (2000) 20289799 10830953 2 (bases 1 to 340000)		
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			

```

* e.mail: hatctori@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
  Beutenbergstrasse 11, D-07745 Jena, Germany,
  * e.mail: gscj-submit@genome.imb-jena.de
  * URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo
  160-8582, Japan,
  * e.mail: nshimizudmb-med.keio.ac.jp
  * URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
  * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
  info.genome@bf.de
  * URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
  * Ihnestrasse 73, D-14195 Berlin, Germany,
  * e.mail: info-chr21@olgen.mpg.de
  * URL: http://chr21.tz-berlin.mpg.de/
  AL163299. Submitted (10-Apr-2000).
  Location/Qualifiers
    1..34000
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="21"
    <1..17637
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="21"
      /map="21q22.3"
      /clone_id="KUD4G11, 5', partial"
      /note="Accession No. AP001060"
      16296..32142
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="21"
      /map="21q22.3"
      /clone_id="KUD6A4"
      /note="Accession No. AP001061"
      25925..62855
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="21"
      /map="21q22.3"
      /clone_id="Q5B10"
      /clone_lib="LI21NCO2-Q Cosmid library"
      /note="Accession No. AP001062"
      56250..101561
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="21"
      /map="21q22.3"
      /clone_id="KUD99F9"
      /clone_lib="KU21-D Cosmid library"
      /note="Accession No. AP001063"
      97304..135091
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="21"
      /map="21q22.3"
      /clone_id="KUD107C4"
      /clone_lib="KU21-D Cosmid library"

```

```

source      /note="Accession No. AP001064"
            109897..225790
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /clone_1ib="Keio BAC library"
            /note="Accession No. AP001065"
            212299..251501
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /clone_1ib="KUD11H9"
            /clone_1ib="KU21-D Cosmid library"
            /note="Accession No. AP001066"
            230165..>340000
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /clone_1ib="KB139C7, 3' partial"
            /clone_1ib="Keio BAC library"
            /note="Accession No. AP001067"
            280806..>340000
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /clone_1ib="RPc11, 3-5 PAC library"
            /note="Accession No. AJ011930"
            1508..13840
            /gene="AIRE"
            join(1508..1759,2178..2352,2599..2754,3138..3212,
            3966..4079,5278..5423,5609..5689,6716..6831,7923..8022,
            8614..8796,9410..9531,10022..10124,12004..12066,
            13277..13840)
            /gene="AIRE"
            /product="autoimmune regulator (APECED protein)"
            /note="Accession No. Z97990"
            1508..1759
            /gene="AIRE"
            /number=1
            1508..1627
            /gene="AIRE"
            join(1628..1759,2178..2352,2599..2754,3138..3212,
            3966..4079,5278..5423,5609..5689,6716..6831,7923..8022,
            8614..8796,9410..9531,10022..10124,12004..12066,
            13277..13348)
            /gene="AIRE"
            /note="Accession No. Z97990"
            /codon_start=1
            /product="autoimmune regulator (APECED protein)"
            /protein_id="BAA95560.1"
            /db_xref="GI:7768776"
            /translation="MATDALRLRLRLHRTETAVVDSAPFLHMLADHDVPEBKFQ
            ETLHLKEKGCPOAFALISWLTQDSTALIDPFWVLFDVNLERYGRLOPLIDSPFK
            DVLSOPRKRPAPVAKLVPPPLPTGRKASEARAPALPGRGASQDLKA
            KPPKPESABOOLPLINGIOTMSASVORAVAMSGDVPAAGAVEGILLQVRESG
            GSKKCIQVGEERTPSKPRDSGSKKARSSGPPRLVAKAGCAAPGCGARLQOQ
            GSVPALPSPDQHKNEDECAVCRDGLICDGCPRAFHLCLSPLEISGTT
            WKCSCLOATVQVQFRAEPPPOBPVETPLPGRLSAGEVRRGPPGFLAGMTTL
            VKHLPAFSAAPLPGIDSSALPLLCVPEGQONLAPARCGVCGDGDVLRCTHL
            AAFHWRCHPACTSRGTGRCRSCGSDVTPAVGVGLAPSPARLAPGAKODTNSHE
            PALHRDLESLSEHTFQDILQWATQSMARPAAPPS"
            2178..2352
            /gene="AIRE"

```

```

exon      /number=2
            2599..2754

Query Match      29.9%; Score 462.4; DB 9; Length 340000;
Best Local Similarity 97.7%; Pred. No. 2.7e-80;
Matches 469; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

183 AGTCGGAGAGACCTCCCTGGGCTGGGCCCCACTGACCTGTGAGAAAGGTTCAATGCG 242
    |||||
6374 AGTGGGAGAGACCTCCCTGGGCTGGGCCCCACTGACCTGTGAGAAAGGTTCAATGCG 6433
    |||||
243 TTGGTACAGTTCGGGGGCCCCCTGGAACGACAGCTGCAAGAAACCGGGTTTCTTC 302
    |||||
303 CCAATAGGATGCCCCCGGGGGGTCTCTTGTGACCAAGATGAGTGGGACAGTGTGT 362
    |||||
6494 CCAATAGGATGCCCCCGGGGGGTCTCTTGTGACCAAGATGAGTGGGACAGTGTGT 6553
    |||||
363 CAGGCGAGATTTCAGGCTTCGACGATGAGAGAGGACAGAGCTGGGAGTTCAAGT 422
    |||||
6554 CAGGCGAGATTTCAGGCTTCGACGATGAGAGAGGACAGAGCTGGGAGTTCAAGT 6613
    |||||
423 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGAAAGAGTGGCTCTCAGAGGGTCTG 482
    |||||
6614 ACCCAGAGATGCTGCTGGGGAGCTGTTTGGAAAGAGTGGCTCTCAGAGGGTCTG 6673
    |||||
483 CACCCAGCCAGTCTGCAATGGGCGCTCTTGTGCTGTCAGAGAAATGAGACAGTGT 542
    |||||
6674 CACCCAGCCAGTCTGCAATGGGCGCTCTTGTGCTGTCAGAGAAATGAGACAGTGT 6733
    |||||
543 GCCGTGTGCGGAGCGGCGGAGCTCATCTGCTGTGACGCGTGGCTTCGCGCTTCAC 602
    |||||
6734 GCCGTGTGCGGAGCGGCGGAGCTCATCTGCTGTGACGCGTGGCTTCGCGCTTCAC 6793
    |||||
603 CTGGCTGCTGCTCTCCCTCCCTCCGTCGCGAGATCCCAAGTGGGACCTGAGGCTTCACG 662
    |||||
6794 CTGGCTGCTGCTCTCCCTCCCTCCGTCGCGAGATCCCAAGTGGGACCTGAGGCTTCACG 6853
    |||||

RESULT 13
AL954247      143192 bp      DNA      linear      PRI 16-OCT-2003
LOCUS
DEFINITION
Pan troglodytes chromosome 22 clone CH251-479113 map 22q22.3,
complete sequence.
ACCESSION
AL954247
VERSION
AL954247.2 GI:37606037
KEYWORDS
HTG.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.
1 (bases 1 to 143192)
REFERENCE
1 The Chimpanzee Chromosome 22 Sequencing Consortium
DNA sequence of chimpanzee chromosome 22 and its evolutionary
implications
Unpublished
2 (bases 1 to 143192)
Kube,M., Sudbrak,R., Mueller,I., Thiel,J., Klages,S., Borzym,K.,
Heltman,K., Gimel,V., Beck,A., Ben Kahla,A., Lehrach,H.,
Yaspo,M.L. and Reinhardt,R.
Direct Submission
Submitted (18-DEC-2002)
The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
*Chinese National Human Genome Center at Shanghai.
Shanghai, China;
*GBF, Dept. of Genome Analysis, Braunschweig, Germany;
*Institute of Molecular Biotechnology, Jena, Germany;
*KRIIB Genome Research Center, Daejeon, Korea;
*Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,

```

Taiwan:

*RIKEN Genomic Sciences Center, Yokohama, Japan.

----- Genome Center

Center: Max-Planck-Institute for Molecular Genetics

Center code: MPIMG

----- Project Information

Center clone name: CH251-479113

----- Summary Statistics

Sequencing vector: pUC18; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 143190 bases at least Q40

Consensus quality: 143192 bases at least Q30

Consensus quality: 143192 bases at least Q20

Quality coverage: 25.4

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. -----

Neighboring sequence information:

This clone is overlapped by

PTB-051003

PTB-061A04

----- SOURCE INFORMATION:

THE CHORI-251 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa, K. et al. Genomics 52:11-8). The DNA from the chimpanzee ('Clint') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Baoji Zhu, Chung Li Shu, Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library characteristics are described at <http://www.chori.org/bacpac/chimpanzee251.htm>. The clone may be obtained from Pieter J. de Jong and coworkers (<http://www.chori.org/bacpac>). VECTOR: pPARAC2.1

IMPORTANT: This sequence is not the entire insert of clone CH251-479113 It may be shorter because we sequence overlapping sections only once, except for a short overlap.

FEATURES

----- Location/Qualifiers

1.143192

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/chromosome="22"

/map="22q22.3"

/clone="CH251-479113"

<1.14090

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="PTB-051003"

/note="overlapping clone"

143189..>143192

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="PTB-061A04"

/note="overlapping clone"

ORIGIN

Query Match 29.2%; Score 451.2; DB 9; Length 143192;

Best Local Similarity 96.2%; Pred. No. 4.7e-78;

Matches 462; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 183 AGTCGGAGAGACCTCCTGGGCTTGAGCCCACTGCTGTGAGAGAGGTTTCATGTGG 242

DB 37910 AGTCGGAGAGACCTCCTGGGCTTGAGCCCACTGCTGTGAGAGAGGTTTCATGTGG 37969

QY 243 TTGCTTACAGTTCCTGGGCCCCCTTGGAAGCAGACCTCTGCAAGAAACCGGTTTCTTC 302

DB 37970 TTGGCTTACAGTTCCTGGGCCCCCTTGGAAGCAGACCTCTGCAAGAAACCGGTTTCTTC 38029

QY 303 CCAATGAGGATGGCCCCCGGGGGGTGCTGTGTGAGACCAAGTGGGAAACAGGTGGT 362

DB 38030 CCAATGAGGATGGCCCCCGGGGGGTGCTGTGTGAGACCAAGTGGGAAACAGGTGGT 38089

QY 363 CAGGGCAGAAATTTACAGCCCTTGCAAGCATGAGGACAGAGCAGAGCTGGGAGTTCAAGT 422

DB 38090 CAGGGCAGAAATTTACAGCCCTTGCAAGCATGAGGACAGAGCAGAGCTGGGAGTTCAAGT 38149

QY 423 ACCCAGAGATGCTGCTGGGGGAGCTGTTTGGGAAAGAGGTGGCTCTCAGAGGGTCTG 482

DB 38150 ACCCAGAGATGCTGCTGGGGGAGCTGTTTGGGAAAGAGGTGGCTCTCAGAGGGTCTG 38209

QY 483 CACCCAGCCAGTGTGATGAGGCGCTCTGCTGTCAGAGAAATGAGACGATGT 542

DB 38210 CACCCAGCCAGTGTGATGAGGCGCTCTGCTGTCAGAGAAATGAGATGT 38269

QY 543 GCCGTGTGTGCGGACGCGCGGAGCTCATCTGCTGTGACGCTGCTTGCGGCTTCCAC 602

DB 38270 GCCGTGTGTGCGGAGCGCGGAGCTCATCTGCTGTGACGCTGCTTGCGGCTTCCAC 38329

QY 603 CTGGCTGCTGCTTCCCTCTCGCTCGGAGATCCCAAGTGGACCTGAGGTCTCCAC 662

DB 38330 CTGGCTGCTGCTTCCCTCTCGCTCGGAGATCCCAAGTGGACCTGAGGTCTCCAC 38389

RESULT 14

AF128772

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

gene

CDS

1..1656

/gene="Aire"

1..>1656

/gene="Aire"

Location/Qualifiers

1.1656

/organism="Mus musculus"

/mol_type="mRNA"

/strain="SJL"

/db_xref="taxon:10090"

/chromosome="10"

/map="between D10Mit31 and D10Mit10"

1..>1656

/gene="Aire"

1..>1656

/gene="Aire"

```
/function="transcription factor"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAF36481.1"
/db_xref="GI:7108573"
/translation="MAGDGMRLRLRLHRTETIAVAIDSAFPLHLADHDVPEDEK
DETLRLKEKEGCPQAFHLLSWLTRDGAIIIDFRRILPKDYNLRYRSLHIDGFP
KVDVINSRKRKPLAGPRAVLPPTPKRALBEPRAVTPATLASKVSQSHLK
TPPKPKPDGNLESQHLPLANGIQTMAASVQRAVTAASGVPTGRGAVEGILLQVFE
GSKKCIQVGESEFTYFNKEDPSGNLKNKASGSLKPVRAKAQVITLPGRDEQV
QOCVPLPLSEPDQVONKDECAVCHDGELICDQCPRAFLIACISPLQETPS
GIMRSCCLOQVQOQLSQPEVSRPELPATPILVGLASAEKTEGSEIKASDA
AVTYNLAHPHAPLALBPSALCPLLSANBERPPAPASRSCVGDGTEVLRCHCA
AAFMHRCHEPTAARPGTNLRCKSCSADSTPPTGPGAVPTSGRPPAPGLAKVDD
ASHDVLHRLDLESLNHSFDGILOMAIQMSRPLAETPPSS"
```

ORIGIN

```
Query Match      25.0%; Score 385.6; DB 10; Length 1656;
Best Local Similarity 68.9%; Pred. No. 5.8e-65;
Matches 580; Conservative 0; Mismatches 244; Indels 18; Gaps 3;

QY 452 TGGGAAAGAGTGGCTCTCAGAGAGGCTGCACCCAGCCCACTGCATGGGCGCTTC
DB 813 TGAGCAGAAAGTGGGCGACAGGTGGGGTCTCTCCCTTCCATCCCTCCAGTAGGCC
QY 512 TTGCCCTGCCGGAAGAATGAGAGAGGTGCGCGTGTGTCGGGAGCGGGGAGCTCAT
DB 873 CCAAGTTAACAGAAACAGAGATGAGTGTGCGGTGCGACAGACGAGGTAGCTCAT
QY 572 CTGTGTGACGGCTGCGCTTCGACCTGACCTGAGCTGCTCCCTCCGCTCCGGAGA
DB 933 CTGTGTGACGGCTGTCCTCCGCGGCTTCACCTGCTGCTGCTCCCACTTCGACGGA
QY 632 GATCCCAATGGGACTGGAGGTGCTCCAGCTGCTGCGACGACACAGTCCAGAGGTGA
DB 993 GATCCCAATGGGCTCTGAGAGTGTCTGCTGCTGCTGCGACGAGGTCCAAAGAACCT
QY 692 GCGCGGGGAGAGAGAGCGCGCGCGCGAGGACCGCGGAGAGCGCGCGTCCCGCGGG
DB 1053 GTTCCAGCTTGAAGGTTCAGAGGCGCGCGAGGCTACCTGACAGAACCGCGATCTCGTGG
QY 752 GCTTAGGTGCGCGGAGAGAGGTGAAGGTTCACCTGAGGAGAACCCCTAGCGCGCATGA
DB 1113 ACTGAGGTGAGCTTTCAGAGAAACAGGGGGCCCATCAGGAGAGCTCAAAGCCAGCTCTGA
QY 812 CAGGACTCTTGTGTACAGACACTGCGCGCTTCGCGACCGCGCTGCGCAGGTCT
DB 1173 TGTGTGTGTCAATATGTGAACCTGTGCGCGCGCACCTGTGACGCTCTCTG-----CT
QY 872 GGACTCTGCGCGCTGACCGCGCTGCTGTGTGTGGGTCTGAGAGGTGACAGAAACCTGGC
DB 1227 GGAAGCTTACAGCACTGTGCGCTTACTGTAGTGTGCGGAATGAGGGCGCGCAGGTCAAG
QY 932 TCTGTGTGCGCGCTTGTGCGGGGTGTGCGAAGATGTGTACGAGCTGTGCGGTGTACTCATG
DB 1287 ACCAAGCGCGCGATGTGAGTGTGTGTGCGCATGTGCACGAGGTGTGCGGTGTGCACACTG
QY 992 CGCGGTGCTTTCATCTGCGCGCTGCGCACTTCCAGCGCGCACTCCCGGCGCGGGAGCGG
DB 1347 TGCCTGTGCTTTCATCTGCGCGCTGCGCACTTCCAGCGCGCGCGCGCGGGAGCAAA
QY 1052 CCGGCGCTGAGATCTCTGCTCAGAGAGCGT-----ACCGCGCGCGCTGTGTGAGGGGGT
DB 1407 TCTCGCTGCAATCTGTCTGTGCAACTGTGACTCCCAAGCGGACACCGGGCGAAGC
QY 1106 GCTGAGCCCCAGCGCGCGCGCTGCGCGCTGCGCA-----GAGTAGCACTGC
DB 1467 TGTACCACTCTGCGCGCGCGCTGCGCACTGCGGCTTGTGCAAGGTAGGGAGCACTCTGC
QY 1160 CAGTACAGAGCGCGCTGTGACAGGATATACCTGTGAGTCTCTTGTGAGGAGACACTT
DB 1527 TAGTACGAGCGCGCTGTCTACATATGAGGAGCAGCTGTGAGTCTCTCTCAATGAGCACTCAT
1586
```

```
QY 1220 CGATGCACTCCGAGTGGGCGCATCCAGAGTAGCGCGCGCGCGCTTCCCTC 1279
DB 1587 TGACGCACTCTGAGTGGGCGCATCCAGAGTAGTACAGCGCGCGCTGCGGACACAC 1646
QY 1280 CT 1281
DB 1647 CT 1648
```

```
RESULT 15
AF128773 1656 bp mRNA linear ROD 26-OCT-2001
LOCUS Mus musculus strain NOD autoimmune regulator (Aire) mRNA, partial
DEFINITION cds.
AF128773
AF128773.1 GI:7108574
```

```
ACCESSION AF128773
VERSION AF128773.1
KEYWORDS Mus musculus (house mouse)
ORGANISM Mus musculus
```

```
REFERENCE 1
AUTHORS Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q.,
Dettler,J.C., Davoodi-Sermitoni,A. and She,J.X.
Chromosomal localization and complete genomic sequence of the
murine autoimmune regulator gene (Aire)
Autoimmunity 31 (1), 47-53 (1999)
20059142
```

```
JOURNAL MEDLINE
PUBMED 10593569
REFERENCE 2 (bases 1 to 1656)
AUTHORS Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q.,
Dettler,J.C., Davoodi-Sermitoni,A. and She,J.X.
Complete genomic sequence, gene structure and localization of the
mouse Aire gene
Unpublished
3 (bases 1 to 1656)
```

```
REFERENCE 3
AUTHORS Ruan,Q.G., Wang C.Y., Shi,J.D. and She,J.X.
Direct Submission
Submitted (17-FEB-1999) Pathology, University of Florida, 1600 SW
Archer Road
```

FEATURES

source

1..1656

/organism="Mus musculus"

/mol_type="mRNA"

/strain="NOD"

/db_xref="taxon:10090"

/chromosome="10"

/map="between D10Mit31 and D10Mit10"

1..>1656

/gene="Aire"

1..>1656

/gene="Aire"

/function="transcription factor"

/codon_start=1

/product="autoimmune regulator"

/protein_id="AAF36482.1"

/db_xref="GI:7108575"

/translation="MAGDGMRLRLRLHRTETIAVAIDSAFPLHLADHDVPEDEK
DETLRLKEKEGCPQAFHLLSWLTRDGAIIIDFRRILPKDYNLRYRSLHIDGFP
KVDVINSRKRKPLAGPRAVLPPTPKRALBEPRAVTPATLASKVSQSHLK
TPPKPKPDGNLESQHLPLANGIQTMAASVQRAVTAASGVPTGRGAVEGILLQVFE
GSKKCIQVGESEFTYFNKEDPSGNLKNKASGSLKPVRAKAQVITLPGRDEQV
QOCVPLPLSEPDQVONKDECAVCHDGELICDQCPRAFLIACISPLQETPS
GIMRSCCLOQVQOQLSQPEVSRPELPATPILVGLASAEKTEGSEIKASDA
AVTYNLAHPHAPLALBPSALCPLLSANBERPPAPASRSCVGDGTEVLRCHCA
AAFMHRCHEPTAARPGTNLRCKSCSADSTPPTGPGAVPTSGRPPAPGLAKVDD
ASHDVLHRLDLESLNHSFDGILOMAIQMSRPLAETPPSS"

ORIGIN

```
Query Match      25.0%; Score 385.6; DB 10; Length 1656;
Best Local Similarity 68.9%; Pred. No. 5.8e-65;
Matches 580; Conservative 0; Mismatches 244; Indels 18; Gaps 3;
```



```

source      1. 1906
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
gene        1. 1906
            /gene="AIRE"
            1. 1659
            /gene="AIRE"
            /function="putative role in transcriptional activation"
            /codon_start=1
            /product="AIRE protein"
            /protein_id="CAB6141.1"
            /db_xref="GI:6706793"
            /db_xref="GDB:Q92083"
            /db_xref="Swiss-Prot:Q92083"
            /translation="MAGDGMRLRLRLHRTETIAVAIDSAFPLHLALADHDVPEDEK
            OETLRLEKEGCPQAFHALLSLTRDSGLIDFMRILFDVYLERSLHSILDFP
            KOVDLNSRKGRPLAGPKAVLPPTPTKRLAEPRATPPATTLASKSVSPSHLK
            TYPKKDNLDSQHPLGNGIQTMASVQRAVTVASGVPRGAVEGILIOVRES
            GSKKCIQVGEFTYTPNKRFBDSGNLKNRSGSLKPVVRAGADVTLIGRDEQKVS
            QOCGVPLSLPSEPVQNKDECAVCHDGGELICDGPRAFHLCSPLOEIPSG
            GLMRSCCLQGRVQOMISQEVSRPELPALERTILVGLSASKEKTRGSPSELASDA
            ATYVNLAPLHPAPLLEPSALCPILSAGNEGRGPAPSRGSCVGDGTEVRLCAACA
            AAFHMRCHPPTAARPTLRCKSGADSTPTPTGTAAPVPSGPRPAPGLAVGDS
            ASHDPVLHNDLSLNLHSFDGILQMAIOSMRPLAETPPSS"
ORIGIN

```

```

Query Match      25.0%; Score 385.6; DB 10; Length 1906;
Best Local Similarity 68.9%; Pred. No. 5.7e-65;
Matches 580; Conservative 0; Mismatches 244; Indels 18; Gaps 3;

QY 452 TGGAGAGAGTGGCTCTCAGAGGGTGTGCACCCCAAGCTCTGCATGGCGCTTC 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 813 TGAGCAGAAAGTGGGCGACAGTGTGGGGTTCTCCCTTCATCCCTCCCAAGTAGCC 872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 512 TTGCTGTGTGACAGATGATGAGACGAGTGTGCGGTGTGCGGACGGCGGAGGCTCAT 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 873 CCAAGTTAACAGAAACAGAGATGATGTGCGGTGTGCGGACGGAGGTAGAGTCTCAT 932
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 572 CTGCTGTGACAGCTGTGCGGCTTCACCTGCGCTGTGCTGCTGCTGCTGCTGCGGGA 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 933 CTGTTGTGACAGCTGTGCGGCTTCACCTGCGCTGTGCTGCTGCTGCTGCTGCGGGA 992
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 632 GATCCCAAGTGGACCTGAGAGTGTCTCAGCTGCTGCGGACCAAGTCCAGAGAGTGA 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 993 GATCCCAAGTGGCTGTGAGAGTGTCTCAGCTGCTGCGGACCAAGTCCAGAGAGTGA 1052
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 GCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1053 GTCCCAAGCTGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 752 GCTTGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1113 ACTGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 812 CAGGATCTTGTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1173 TGTGTGTGTACATATGTGTAACTGTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 872 GGAAGTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1227 GGAAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 932 TCTGTGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 991
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1287 ACCAAGCGGCGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 992 CCGCGGCTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1051
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1347 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1052 CTTGCGCTGTGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 1407 TCTCCGCTGCAATCTCTCTGTGACAGATCTGACCTCCAGCGCACACCGGCGAAGC 1466
QY 1106 GCTGCGCCCGCAGGCGCCGCGCGCTGCGCCCTGCGGAGCTGCGCA-----GGATGACACTGC 1159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1467 TGTACCACTCTGTGAGCGCCCGCTGTGACAGCTGTGCGGAGTGTGAGGAGACAGCTGTGC 1526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1160 CAGTACAGAGCGGCTGTGTGACAGAGAGATGAGAGTCTCTTCTGAGGAGACAGACTT 1219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1527 TAGTACAGAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1220 CGATGCGATCTGTGAGTGGGCGCATCCAGAGAGATGAGGCGCGCGCGCGCGCGCGCTTC 1279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1587 TGAGCGCATCTGTGAGTGGGCGCATCCAGAGATGTACAGCGCGCGCTGTGCGGAGACAGCACC 1646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1280 CT 1281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1647 CT 1648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 19
AF128116      1924 bp      mRNA      linear      ROD 29-FEB-2000
LOCUS
DEFINITION
Mus musculus autoimmune regulator (Aire) mRNA, alternatively
spliced product 1c, complete cds.
AF128116
AF128116.1 GI:7108533
VERSION
KEYWORDS
SOURCE
MUS MUSCULUS (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1924)
Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
Expression and alternative splicing of the mouse autoimmune
regulator gene (Aire)
Unpublished
2 (bases 1 to 1924)
Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
Direct Submision
Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW
Archer Road, Gainesville, FL 32610, USA
Location/Qualifiers
1. 1924
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1. 1924
/gene="Aire"
58..1704
/gene="Aire"
/function="transcription factor"
/notes="alternatively spliced product 1c"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAF36461.1"
/db_xref="GI:7108534"
/translation="MAGDGMRLRLRLHRTETIAVAIDSAFPLHLALADHDVPEDEK
OETLRLEKEGCPQAFHALLSLTRDSGLIDFMRILFDVYLERSLHSILDFP
KOVDLNSRKGRPLAGPKAVLPPTPTKRLAEPRATPPATTLASKSVSPSHLK
TYPKKDNLDSQHPLGNGIQTMASVQRAVTVASGVPRGAVEGILIOVRES
GSKKCIQVGEFTYTPNKRFBDSGNLKNRSGSLKPVVRAGADVTLIGRDEQKVS
QOCGVPLSLPSEPVQNKDECAVCHDGGELICDGPRAFHLCSPLOEIPSG
GLMRSCCLQGRVQOMISQEVSRPELPALERTILVGLSASKEKTRGSPSELASDA
ATYVNLAPLHPAPLLEPSALCPILSAGNEGRGPAPSRGSCVGDGTEVRLCAACA
AAFHMRCHPPTAARPTLRCKSGADSTPTPTGTAAPVPSGPRPAPGLAVGDSASHD
PVLHNDLSLNLHSFDGILQMAIOSMRPLAETPPSS"
ORIGIN
Query Match      25.0%; Score 385.6; DB 10; Length 1924;
Best Local Similarity 68.9%; Pred. No. 5.7e-65;

```

Matches	580; Conservative	0; Mismatches	244; Indels	18; Gaps	3;
QY	452 TGGGAAGAGTGGCTCTCAGAGAGGTGTGCAACCCAGCCCAAGTCTGATGGGCGTCTC				511
Db	858 TGAGCAGAAAGTGGGCGACAGTGGGGTCTCCCTCCATCCATCCCTCCAGATGAGCC				917
QY	512 TTGCCGTGCCAAGAATGAGAGAGAGTGTGCGGTGTGTCGGGAAGGGGGAGTCAAT				571
Db	918 CCAAGTTAACAGAAAGACAGAGATGATGTGCGGTGTGTCAGACAGAGAGTGAAGTCAAT				977
QY	572 CTGCTGTGACGAGTGGCTCTCGAGCCCTTCCACCTGGCTGCTGTCCCTCCGCTCCGGA				631
Db	978 CTGTTGTGACGAGTGTGCCCCGGGCTTCCACCTGGCTGCTGTCCCACTCTGACAGA				1037
QY	632 GATCCCATTTGGGACCTGAGAGTGTCTCCAGCTGCTCTGACGAGCAAGTCCAGAGAGTGA				691
Db	1038 GATCCCAATGTGCTCTTGAGATGTCTCCGCTGCTCCAGGGGAGAGTCCACAGAACTT				1097
QY	692 GCCCGGGGAGAGAGACCCCGGCCCCAGAGACCCGCTGAGAACCCCGCTCCCCCGGG				751
Db	1098 GTCCACGCTTGAGTGTCCAGGCCCCCGAGCTACTGACAGAGACCCCGATCTCGTGG				1157
QY	752 GCTTAGTGTGGGGAGAGAGATGAAGTCCACCTGGGGAACCCCTAGCCGGCATGGA				811
Db	1158 ACTGAGGTGAGCTTCAAGAAACACAGGGGCCATCCAGGAGACTCAAGCCAGCTCTGA				1217
QY	812 CACGACTCTTGTCTACAGACCTGCGGCTCCGCTTGTGACGCCCGCTGCCAGTCT				871
Db	1218 TGTGTGTACATAATGTAACCTGTGCGCCCGACCCCTGACGCTCTCTG-----CT				1271
QY	872 GGACTCTCGGCTGACACCCCTACTGTGTGTGGTCTCTGAGGGTCCAGACAGAACTGGC				931
Db	1272 GGAAGCTTACGACTGTGCTCTCTACTAGTGTCTGGGAATGAGGGCGGCGAGTCCAGC				1331
QY	932 TCTGTGTGCGCTTGTGCGGAGTGTGCGAAGTGTACGACGTGTCTCGGTGTACTGACTG				991
Db	1332 ACCAAGCGGCGCATGTGATGTGTGGCATGGCAACGAGTGTTCGGGTGTGCAACTG				1391
QY	992 CGCGGTGCTTCCACTGGCGCTGACACTTCCAGCGCGGACCTCCGCGCGGGGAGCGGG				1051
Db	1392 TGCCTGCTTCCACTGTGCGCTGACACTTCCGACGCGCGCGCGCGGGGAGCAAA				1451
QY	1052 CCTGCGCTGACATCTGTCTCAGAGACGTG-----ACCCGACCCCTGTGTGAGGGGT				1105
Db	1452 TCTCGCTGCAATCTGTCTGTGCAACTGCACTCCAGCGACAGACACCGGCGCAAGC				1511
QY	1106 GCTGCGCCCAAGCCCGCGCTGCGCTTGGGCTTGGGCTTGGCA-----GATGACACTGC				1159
Db	1512 TGTACCACTCTGTGGCGCGCTGCAACACTGGCTTGGCCAAAGTGTAGGGGACGACTGTGC				1571
QY	1160 CAGTACAGAGCGCGCTGTGACAGAGGATGACCTGTGAGTCCCTTCTGAGGAGACACTT				1219
Db	1572 TGTGTGACACCTGTTCTCATATGAGGACAGACTGTGCTCTCTCATATGAGCACTCAT				1631
QY	1220 CGATGCACTCTGTGACGTGGGCACTCCAGAGACTGCGCTGCGCGGCGCCCTTCCCTTC				1279
Db	1632 TGAAGGCACTGTGCAAGTGGGCACTCCAGAGCAATGTACCGCGGCTGCGCGAGACACACC				1691
QY	1280 CT 1281				
Db	1692 CT 1693				

RESULT 20

AF079536 1936 bp mRNA linear ROD 17-MAR-1999
 DEFINITION Mus musculus autoimmune regulator (Aire) mRNA, complete cds.
 ACCESSION AF079536
 VERSION AF079536.1 GI:4426598
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (baes 1 to 1936)
TITLE	Wang,C.Y., Shi,J.D., Davoodi-Semiromi,A. and She,J.X. Cloning of Aire, the mouse homologue of the autoimmune regulator (AIRE) gene responsible for autoimmune polyglandular syndrome type 1 (ASPI)
JOURNAL	Genomics 55 (3), 322-326 (1999)
MEDLINE	99168902
PUBMED	10049587
REFERENCE	2 (baes 1 to 1936)
AUTHORS	Wang,C.Y., Shi,J.D., Davoodi-Semiromi,A. and She,J.X.
TITLE	Direct Submission
JOURNAL	Submitted (21-JUN-1998) Pathology, Immunology and Laboratory Medicine, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA
FEATURES	Location/Qualifiers
source	1..1936
gene	/organism="Mus musculus"
CDS	/mol_type="mRNA"
	/strain="B6"
	/db_xref="taxon:10090"
	/cissue_type="thymus"
	1..1936
	/gene="Aire"
	58..1716
	/gene="Aire"
	/function="transcription factor"
	/codon_start=1
	/product="autoimmune regulator"
	/protein_id="AAD20444.1"
	/db_xref="GI:4426599"
	/translation="MAGDGMRLRLRLHRTETAVAIADSAFPLHALADHDVVPBPKI OETLAKKRGCPDPAFHALLLSWLTRDGAALIDFWRILLKYNLERYSRLSLIDGFL KDVLINQSRKGRKPLAAGKAYLPPRPPTKRALEPPATPTLASVSRGSHL TKPRKPDGNLSQHLPIGNGIQTMAASVQRAVTAAGVPEPRGAQVITPGDEKVG GSKKCIQVGEFYPYPNKFEDEBSGMLKKARSGSLKVPRAKQAQVITPGDEKVG QCGVPLPSLSEPVQNKDEBCAVCHDGSEILCCGCPAPFAHACLSPLQETPS GLMRCSCICGQVOONLSQPEVSRPELPAETPIVLAGRSASERTGRSELKASDA AVTYNLLAPHPAFLPALSALCPLLSAGNEBRPGAPASRCVCGDGEVLRCAHCA AAFHNRCHFPPTAARPTNLRKSCSADSTPTPTGPAVPTSGRPAFLAKVGDSD ASHDEVLRHDDLESILNHSFDGIILQMAIQMSRPLAETPPSS"
ORIGIN	
Query Match	25.0%; Score 385.6; DB 10; Length 1936;
Best Local Similarity	68.9%; Pred. No. 5.7e-65;
Matches	580; Conservative 0; Mismatches 244; Indels 18; Gaps 3
QY	452 TGGGAAGAGTGGCTCTCAGAGAGGTGTGCAACCCAGCCCAAGTCTGATGGGCGTCTC 511
Db	870 TGAGCAGAAAGTGGGCGACAGCACTGTGGGTTCTCCCTCCATCCCTCCAGATGAGCC 929
QY	512 TTGCCGTGCCAAGAATGAGAGAGTGTGCGGTGTGTCGGGAAGGGGGAGTCAAT 571
Db	930 CCAAGTTAACAGAAAGACAGAGATGATGTGCGGTGTGTCAGACGAGAGTGAAGTCAAT 989
QY	572 CTGCTGTGACGAGTGGCTCTCGAGCCCTTCCACCTGGCTGCTGTCCCTCCGCTCCGGA 631
Db	990 CTGTTGTGACGAGTGTGCCCCGGGCTTCCACCTGGCTGCTGTCCCACTCTGACAGA 1044
QY	632 GATCCCAATGTGACCTGTGAGGTGTCTCAGCTGTCTGACGSCAACAGTCCAGAGGTGA 691
Db	1050 GATCCCAATGTGCTCTTGAGATGTCTCTGCTGCTCCAGGCGAGAGTCAACAGAACTT 1109
QY	692 GCCCGGGGAGAGAGACCCCGGCCCCAGAGACACCCGTGAGAACCCCGCTCCCCCGGG 751
Db	1110 GTCCACGCTTGAGTGTCCAGGCCCCCGAGCTTACTGTGACAGACCCCGATCTCGTGG 116
QY	752 GCTTAGTGTGGGGAGAGAGATGAAGTCCACTGGGGAACCCCTTAGCCGGCATGGA 811
Db	1170 ACTGAGTGTAGCTTCAAGAAACACAGGGGCCCATTCAGAGGAGCTCAAGCCAGCTCTGA 122
QY	812 CACGACTCTTGTCTACAGACCTGCGGCTCCGCTTGTGACGCCCGCTGCCAGTCTC 871

Db 1230 TGTGCTGTCAATATGTGAACCTGCTGGCCCGCACTGACCTCTCTG-----CT 1283
Qy 872 GGAATCTCTGGCCCTTCAACCCCTACTGTGTGTGGTCTTGAAGGATCAAGAACCTGGC 931
Db 1284 GGAGCTTCAAGCACTGTGCTTACTAGAGTGTGGGAATGAGGGGCGGCAAGTCCAGC 1343
Qy 932 TCTGTGTGGGCTGTGGGGGTGTGGGGAATGTGTACGGAAGTGTGTGGGTACTAG 991
Db 1344 ACCAAGCGGCGGAGT 1403
Qy 992 CGCGCTGTGCTTCACTGTGGGCTGTGCACTTCCAGCGGCACTCCCGGCGGGAAG 1051
Db 1404 TGTGCTGTGCTTCACTGTGGGCTGTGCACTTCCAGCGGCGGCGGCGGGAAG 1463
Qy 1052 CTGTGCTGTGCAATCTGTCTGTGAGAGAGCTG-----ACCCAGCCCTGTGTGAGGGGCT 1105
Db 1464 TGTGCTGTGCAATCTGTCTGTGAGAGAGCTGCACTGCCAGCGGAGGCAACCGGCGGAAG 1523
Qy 1106 GGTGGCCCGCAAGCCCGCCCGCTGTGGCCCTGTGGCCCA-----GGATGACATG 1159
Db 1524 TGTACCACTCTGTGGCCCGCTGTGAGAGAGCTGTGGCCAGGTAGGGGAGCACTGTG 1583
Qy 1160 CAGTCAAGAGCCGCTGTGCAAGAGATGACTGTAGTCCCTTGTGAGGAGAGCACTT 1219
Db 1584 TGTGCAAGAGCCGCTGTGCAAGAGATGACTGTAGTCCCTTGTGAGGAGAGCACTT 1643
Qy 1220 CGATGCAATCTGTGAGTGGGCGATCAAGAGATGGCCGCTGTGGCGGCGGCTTCCCTTC 1279
Db 1644 TGAAGGCAATCTGTGAGTGGGCGATCAAGAGATGATGACGCGCTGTGGCGGAGACACAC 1703
Qy 1280 CT 1281
Db 1704 CT 1705

RESULT 21
AF128117 1921 bp mRNA linear ROD 29-FEB-2000
LOCUS AF128117
DEFINITION Mus musculus autoimmune regulator (Aire) mRNA, alternatively
spliced product id, complete cds.
ACCESSION AF128117.1 GI:7108535
VERSION AF128117.1 GI:7108535
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
MUS musculus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
Expression and alternative splicing of the mouse autoimmune
regulator gene (Aire)
Unpublished
2 (bases 1 to 1921)
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
Direct Submission
Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW
Archer Road, Gainesville, FL 32610, USA
FEATURES
source
1. .1921
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1. .1921
/gene="Aire"
58. .1701
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product id"
/codon_start=1
/product="autoimmune regulator"

ORIGIN

Query Match 24.3%; Score 375.6; DB 10; Length 1921;
Best Local Similarity 71.1%; Pred. No. 5e-63;
Matches 546; Conservative 0; Mismatches 204; Indels 18; Gaps 3;

/protein_id="AA036462.1"
/db_xref="GI:7108535"
/translation="MAGDGMRLRLRLHRTETIAVAIDAPFLLAADHDVPEKDF
DETLRLKEKGGPOAFHALLSWLITRDEPAIIDLPMILDKDNLREYSLHSLIDGP
KQVDLNQSGKGLAGPKAAVLPRPPKRAALAEPRPAPATLAKSVSSSHLK
TKPPKPDGNTLSOHLPLNGIOTMAASVQRAVYAGVPGTRGAVEGILIQOVES
GSKKIOYGGFPYTPNKREDSGLKXKASGSSLKPVRAKAGORPEOVYGOOC
VPPPLSLEPQVONQNEBCAVCHNGBELICDDGEPRAHLALCLPPEIOTSGLMK
SCCLQKRVQNLSPALCPILSAGNEGRPEAPASRCSVGDGTBLRCAACAAAFHW
NLIAPHPAPLLEPSPALCPILSAGNEGRPEAPASRCSVGDGTBLRCAACAAAFHW
RCHFPAPARPTNLRKCSADSTPTPTPEAVTSGTPAPGLAKVGDSDASHP
VLRHDLLESLMHSFPGILQWALQMSRPLAETPPFS"

Qy 526 AGAATGAGAGAGATGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 585
Db 929 AGAATGAGAGATGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 988
Qy 586 GCGCTGGGCTTTCACATGAGCTGTGCTGCTCCCTCGGAGATGCCAGTGGGA 645
Db 989 GTTCCCGGGCTTTCACATGAGCTGTGCTGCTCCCTCGGAGATGCCAGTGGG 1048
Qy 646 CTTGAGAGTCTTCCAGCTGCTGTGAGGCAACAGTCCAGAGGTGACGCCGCGGAGAG 705
Db 1049 TGTGAGAGATGCTTCTGCTGCTGTGAGGAGAGAGTCCAGACAACTGTCCAGCTGAG 1108
Qy 706 AGCCCGGCGGCGAG 765
Db 1109 TGTGAG 1168
Qy 766 GAGAGAGAGTAAAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825
Db 1169 CAG 1228
Qy 826 ACAAG 885
Db 1229 ATGTGAG 1282
Qy 886 TGAAG 945
Db 1283 TGTGCTTCTACTGAGT 1342
Qy 946 GCGGGGTGTGCGAGATGT 1005
Db 1343 GCAAGT 1402
Qy 1006 ACTGAGGCTGTGCACTTCCAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1065
Db 1403 ACTGAGGCTGTGCACTTCCAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1462
Qy 1066 CTTGTGAG 1119
Db 1463 CTTGTGAG 1522
Qy 1120 CGGCGGCTGTGAG 1173
Db 1523 GCGCCGCTGTGAG 1582
Qy 1174 CTGTGAG 1233
Db 1583 TTCTACATGTGAG 1642
Qy 1234 AGTGGGCAATCAAG 1281
Db 1643 AGTGGGCAATCAAG 1690

RESULT 22
AF128115

LOCUS AF128115 1933 bp mRNA linear ROD 29-FEB-2000
DEFINITION Mus musculus autoimmune regulator (Aire) mRNA, alternatively
spliced product 1b, complete cds.
ACCESSION AF128115
VERSION AF128115.1 GI:7108531
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1933)
Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
Expression and alternative splicing of the mouse autoimmune
regulator gene (Aire)
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1933)
Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
DIRECT SUBMISSION Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW
JOURNAL Archer Road, Gainesville, FL 32610, USA
FEATURES
source 1..1933
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1933
/gene="Aire"
58..1713
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 1b"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AF36460.1"
/db_xref="GI:7108532"
/translation="MAGDQMLRRLRLRLTEIAVAIDSAFPLHLADHDVPEDKF
OETLRKEKGGCPAFAHLSLITLTPDGLIDPMLIPKDYVLEYSRSLIDP
KDYDLNOSKRKRPPLAGPKAVLPPTPKRLKLEPRATPPTTLASKSVSPGSHLK
TPPKPNDNLSEQLPLANGIQTMAVQRAVTASGVPGTGAVESILIQVYES
GSKKCIQVGEFYTNPKNEDPBGINKNARSSSKPVVRAGAQTTPGRDQVYG
QCGVPLPLSPSEPVNONDECAVCHDGSELCIDCGEPRAHLACLPLOEIPSG
LWRSCCLGRRVQONLSQPSRSPPLPATPTLVGLRSSEKTRSPSLKASDA
VTVYVNLAPHPAPLLEPSALCPILSAGNEGPRGPASARCSYCGCTGVLCAHDA
AHRMRCHPTTAARPPTNLRCKSCSDSTPTPTPEAVPTSGPRAPGLAKVGDSDA
SHDPVLRDLESLINEHSPDGLIQWAIQSMSPLAETPPFS"

ORIGIN
Query Match 24.3%; Score 375.6; DB 10; Length 1933;
Best Local Similarity 71.1%; Pred. No. Se-63;
Matches 546; Conservative 0; Mismatches 204; Indels 18; Gaps 3;
QY 526 AGAATGAGACGAGTGTGCGGAGCGGAGGAGTCACTGTGTGACGGCT 585
DB 941 AGAACAGAGTAGTGTGCGGAGCGGAGGAGTCACTGTGTGAGGGCT 1000
QY 586 GCCCTGGGCGCTTCACCTGAGCTGCTGCTCCGCTCCGGAGATCCCAATGGGA 645
DB 1001 GTCCCGGGGCTTCACCTGAGCTGCTGCTCCGCTCCGGAGATCCCAATGGGA 1060
QY 646 CCTGAGAGTCTCAGAGCTCAGAGCAACAGTCCAGAGAGTGAAGAGAGAGAG 705
DB 1061 TCTGAGAGTCTCAGAGCTCAGAGCAACAGTCCAGAGAGTGAAGAGAGAG 1120
QY 706 AGCCCGGCGCCGAGAGCAACCCGCTGAGAGAGAGAGAGAGAGAGAGAG 765
DB 1121 TGTCCAGAGCGCCGAGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAG 1180
QY 766 GAGAGAGAGTGAAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 825
DB 1181 CAG 1240

QY 826 ACAGAGACCTGCGGCTCCGCTTCTGAGAGCGCCGCTGCAAGTCTGCTCGGCC 885
DB 1241 ATGTGAACCTGTGGGCTCCGACCTTGACAGTCTCTCTG-----CTGAGCTTACAGCAC 1294
QY 886 TGCACCCCTACTGTGTGTGGGTCTGAGAGTCAAGCAACCTGACTCTGTGCGCTT 945
DB 1295 TGTGCTCTTACTGAGTGTGTGGGAATGAGAGGCGGCGAGTCCAGACCAAGCGCGCAT 1354
QY 946 GCGGGGTGTGCGAAGTGTACGAGAGTCTGCGGTACTACATGCGCGCTGCTTTC 1005
DB 1355 GGAAGTGTGTGCGAAGTGTACGAGAGTCTGCGGTACTACATGCGCGCTGCTTTC 1414
QY 1006 ACTGAGCGTCCACTTCCAGCGGAGCACTCCGAGCGCGAGAGAGAGAGAGAGAG 1065
DB 1415 ACTGAGCGTCCACTTCCAGCGGAGCACTCCGAGCGCGAGAGAGAGAGAGAGAG 1474
QY 1066 CTGTCTCAGAGACGTG-----ACCCAGCCCTGTGTGAGGGGTGTGAGCCCGAGCC 1119
DB 1475 CCGTCTGTGCAAGTCTGACTCCAGCGGAGCAACCGGCGAGAGTGTACCACTCTG 1534
QY 1120 CCGCCCGCTGCGCCCTGTGAGGCTTGCCTCA-----GATGACACTGCGAGAGCCG 1173
DB 1535 GGCCTGCTGCAAGTCTGACTCCAGCGGAGCACTCCGAGCGCGAGAGAGAGAG 1594
QY 1174 CTCTGACAGAGAGATGACGTGAGTCCCTTCTGAGAGAGACACCTTCATGAGCATCTTGC 1233
DB 1595 TTCTACTATGAGAGCAAGTCTGAGTCTCTCTCAATGAGCATCTTATGAGCATCTTGC 1654
QY 1234 AGTGGCCATCCAGAGCATGAGCGGCTCGCGGCGCCCTTCCCTCT 1281
DB 1655 AGTGGCCATCCAGAGCATGAGCGGCTCGCGGCGCCCTTCCCTCT 1702

RESULT 23
HSLTPEK 2586 bp DNA linear PRI 29-APR-1996
LOCUS HSLTPEK
DEFINITION H.sapiens gene for liver type phosphofructokinase.
ACCESSION X80853
VERSION X80853.1 GI:1292864
KEYWORDS phosphofructokinase.
Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1
Levanon,D., Brandeis,M., Bernstein,Y. and Groner,Y.
Common promoter features in human and mouse liver type
phosphofructokinase gene
JOURNAL Biochem. Mol. Biol. Int. 35 (5), 929-936 (1995)
MEDLINE 96018039
PUBMED 7549935
REFERENCE 2 (bases 1 to 2586)
Groner,Y.
DIRECT SUBMISSION Submitted (04-AUG-1994) Y. Groner, The Weizmann Institute, Rehovot
JOURNAL 76100, ISRAEL
FEATURES
source 1..2586
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_1fb="chromosome 21 enriched genomic library
(cytometry 7;411-417;1986)"
1..2586
/product="phosphofructokinase"
/note="liver type"
447..616
/note="1/2Alu"
/tpc_family="Alu"
1251..1597
repeat_region
repeat_region
repeat_region

```
misc_binding      /rpt_family="Alu"
                  2471. .2526
ORIGIN            /bound_motety="Sp1"
```

Query Match	21.7%	Score 336;	DB 9;	Length 2586;
Best Local Similarity	100.0%	Pred. No. 2.5e-55;		
Matches 336;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	1210	AGCAACCTTGGATGGCAATCCGTGACAGTGGGCGCATCAAGAGCATGGCCGCTCGGCGGCC	126
Db	135	AGCAACCTTGGATGGCAATCCGTGACAGTGGGCGCATCAAGAGCATGGCCGCTCGGCGGCC	194
OY	1270	CCTTCCCTCCTGACCCCGAGATGGCCGGGACATGCACTCTGATGAGAGTGCTGAGAA	1322
Db	195	CCCTCCCTCCTGACCCCGAGATGGCCGGGACATGCACTCTGATGAGAGTGCTGAGAA	254
OY	1330	GGACACCTCCTTCCTCAGTCCGTGGAAGCGCGCCGCTGGGATCAAGAAAGGGACAGCCG	1385
Db	255	GGACACCTCCTTCCCTCAGTCCGTGGAAGCGCGCCGCTGGGATCAAGAAAGGGACAGCCG	314
OY	1390	ACCTTTGTCAATGCTCGGCTGTAAACAGCTCTGTGTTCTTGGGGACACCAAGCATCATG	1442
Db	315	ACCTTTGTCAATGCTCGGCTGTAAACAGCTCTGTGTTCTTGGGGACACCAAGCATCATG	374
OY	1450	TGCGTGAATTTAAACCTGCGCCCACTTCTCACTCTGGAGTCCCGGGAGCCCTCTCCT	1505
Db	375	TGCGTGAATTTAAACCTGCGCCCACTTCTCACTCTGGAGTCCCGGGAGCCCTCTCCT	434
OY	1510	TGCGTGTGACCTACTATAAAATATATAAATTAAGCTG	1545
Db	435	TGCGTGTGACCTACTATAAAATATATAAATTAAGCTG	470

ORIGIN

Query Match	11.3%;	Score 174.6;	DB 10;	Length 1744;
Best Local Similarity	69.9%;	Pred. No. 7.1e-24;		
Matches 269;	Conservative 0;	Mismatches 104;	Indels 12;	Gaps 2;

Oy	909	CTTAGGGGTACGAGAAACCTGGCTTCCTGGGTGCGGGTGGCGGGATGTGCGAGATAGGTACG	968
Db	1129	CTTCAGAGAGACCCGGGTCCAGCACCAAGCGCGGATGCAATGTGTGTGGCGATGGCAC	1188
Oy	969	GACGTGCTGCGGTGTACTACTGCGCGGCTGCTTTCACATGGACGCTGCCACTTCCAGCC	1028
Db	1189	GAGGTGTGGGGTGTGCACACTGTGGCGGTGCTTTCACATGGCGCTGCCACTTCCGACG	1248
Oy	1029	GGCACCTTCGGGCCCCGGGACGGGCTGTGGCTGCGAGATCTGTCTCAGAGACGTG-----	1082
Db	1249	GCCCCCCCCCGGGCGGGAGCAATCTCCGTGCAGAACTCTGCTGCGAGATCGACCTCC	1308
Oy	1083	ACCCCAAGCCCTGTGTGAGGGGGGTCTGGGCCCCGACCCCGCGCGCTGGGCCCT	1142
Db	1309	ACGCGAGGACACCCGGGGGAGAGCTGTACCCACTCTGGGCCCCGTCACAGCACTGGGCTT	1366
Oy	1143	GCCAA-----GGATGACACTGTCACAGTACAGAGCCGCTCTGCA CAGGATGACTGGAG	1196
Db	1369	GCCAAAGTAAAGGGAGACGACTGTGCTAAGTACAGACCCTGTTCACATAAGGAGAGCACTGGAG	1422
Oy	1197	TCCCTTCTGAGCGAGCAACACTTGTGATGGCATCTTGCAGTGGGCCATTCAGAGCATGGCC	1252
Db	1429	TCCCTCTCTCAATGAGCACTCATTTGAGCGGCACTCTGAGTGGGCCATTCAGAGCATGTGCA	1488
Oy	1257	CGTCCGGGCGGCCCTTCCCTCTCT	1281
Db	1489	CGCCCGCTGGCCGAGACCACTCT	1513

FEATURES	
source	location/Qualifiers
gene	1..1744
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="B6"
	/db_xref="taxon:10090"
	/chromosome="10"
	/map="between D10Mit31 and D10Mit10"
	1..1744
	/gene="Aire"
CDS	58..1524
	/gene="Aire"
	/function="transcription factor"
	/note="alternatively spliced product 2d"
	/codon_start=1
	/product="autoimmune regulator"
	/protein_id="AF36466.1"

```

RESULT 25
AF128120
LOCUS
DEFINITION
Mus musculus autoimmune regulator (Aire) mRNA, alternatively
spliced product 2c, complete cds.
AF128120
ACCESSION
AF128120.1 GI:7108541
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1747)
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
Expression and alternative splicing of the mouse autoimmune
regulator gene (Aire)
Unpublished
2 (bases 1 to 1747)
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
Direct Submission
Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW
Archer Road, Gainesville, FL 32610, USA
JOURNAL
TITLE
FEATURES
source
1..1747
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6"
/db_xref="taxon:10090"

```


VERSION AF128118.1 GI:7108537
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1759)
AUTHORS Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
TITLE Expression and alternative splicing of the mouse autoimmune regulator gene (Aire)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1759)
AUTHORS Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
TITLE Direct Submision
JOURNAL Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA
FEATURES
source
1..1759
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1759
/gene="Aire"
58..1539
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 2a"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAF36463.1"
/translation="MAGDGMRLRLRLHRTETIAVAIDSAFPLHLADHDVPEDEK
OETLRKEGEGPOAFHALLSWLITRDSGAILDFWRIIFKDYNLERYSGLSILDFP
KVDLNSRKGKPLAGPRAVLPPEPRKRALEPRATPATLASKSVSPGSLIK
TKPKKPDGNLESQHLPLNGIOTMAASVORAVTASGVPTGAVESILILQOYES
GRSKKCIQVGEFTYFNKFEDEPSGNLKNKASGSLKPVNRAKAGQVITPGDEQVGS
QOCQVPLPSPEPQVKNDECAVCHDGBELICDDCPRAFLACLSPLQETPS
AFMRSCCICQVQNLISQPEVSRPELPALPTGPAISARCSVCGDTEVLCAHCAA
AFMRCHPPTAARPGTINLRCSKSDSPTEGTGPEAVTSGPRAPFELATKVGDDSA
SHDPLVLRHDLSESLNHSFDDILQALISMRPLAETPPFS"

gene
1..1759
/gene="Aire"
58..1539
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 2a"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAF36463.1"
/translation="MAGDGMRLRLRLHRTETIAVAIDSAFPLHLADHDVPEDEK
OETLRKEGEGPOAFHALLSWLITRDSGAILDFWRIIFKDYNLERYSGLSILDFP
KVDLNSRKGKPLAGPRAVLPPEPRKRALEPRATPATLASKSVSPGSLIK
TKPKKPDGNLESQHLPLNGIOTMAASVORAVTASGVPTGAVESILILQOYES
GRSKKCIQVGEFTYFNKFEDEPSGNLKNKASGSLKPVNRAKAGQVITPGDEQVGS
QOCQVPLPSPEPQVKNDECAVCHDGBELICDDCPRAFLACLSPLQETPS
AFMRSCCICQVQNLISQPEVSRPELPALPTGPAISARCSVCGDTEVLCAHCAA
AFMRCHPPTAARPGTINLRCSKSDSPTEGTGPEAVTSGPRAPFELATKVGDDSA
SHDPLVLRHDLSESLNHSFDDILQALISMRPLAETPPFS"

ORIGIN

Query Match 11.3%; Score 174.6; DB 10; Length 1759;
Best Local Similarity 69.9%; Pred. No. 7.1e-24;
Matches 269; Conservative 0; Mismatches 104; Indels 12; Gaps 2;

QY 909 CCTGAGGCTCAGACAACTGGCTCTGTCGCGCGGTGGGGGTGGCGAGATGCTACG 968
DB 1144 CCTGACGAGACCCCGGCTCAGACCAACGCGCGATGACGTGTGTGGCGGATGCGACCC 1203
QY 969 GACGTGCTGCGGTGTACTACTGCGCGGCTCTTCACTGGGCGTGCACCTCCAGCC 1028
DB 1204 GAGGTGTTCGGGTGTCAACTGTGCGCTGCTTCACTGGGCGTGCACCTCCAGCG 1263
QY 1029 GGCACCTCCCGGCGCGGACGCGGCTGCGCTGCAGATCTGTCTCAGAGACGTG----- 1082
DB 1264 GCGCGCGCGCGCGCGGACCAATCTCCGCTGCAAACTGTGCTGCAAGCTGACTGCC 1323
QY 1083 ACCCGCGCGCTGTGAGAGGGGTGTGCGGCCCGCGCGCGCGCGCTGCGCGCTTGGGCGCT 1142
DB 1324 AGCGCCAGGACACCGGCGCGCAAGCTGTACCACTTGGGCGCGCTGCAAGCTTGGGCGCTT 1383
QY 1143 GCCAA-----GGATGACACTGCGCAGTACAGAGCCGCTGTGCAAGGAGTGACTGGAG 1196
DB 1384 GCCAAGTAGGGAGCACTGTCTTACTACGACCTCTTCTTACATAGGAGCACTGGAG 1443
QY 1197 TCCCTTCTGAGCAGACACTTTCATGAGCATCTGCACTGAGGCGCATCCAGAGCATGGCC 1256
DB 1444 TCCCTCTCAATGAGCATCTTTCATGAGCATCTGCACTGAGGCGCATCCAGAGCATGTCA 1503

QY 1257 CGTCCGCGGCGCGCGCGCTTCCCTCTCT 1281
DB 1504 CGCCGCTGCGCGGACACACACCTT 1528

RESULT 28
AF128124
LOCUS 1625 bp mRNA linear ROD 29-FEB-2000
DEFINITION Mus musculus autoimmune regulator (Aire) mRNA, alternatively
spliced product 3c, complete cds.
ACCESSION AF128124
VERSION AF128124.1 GI:7108549
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1625)
AUTHORS Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
TITLE Expression and alternative splicing of the mouse autoimmune regulator gene (Aire)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1625)
AUTHORS Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
TITLE Direct Submision
JOURNAL Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA
FEATURES
source
1..1625
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1625
/gene="Aire"
58..1275
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 3c"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAF36469.1"
/translation="MAGDGMRLRLRLHRTETIAVAIDSAFPLHLADHDVPEDEK
OETLRKEGEGPOAFHALLSWLITRDSGAILDFWRIIFKDYNLERYSGLSILDFP
KVDLNSRKGKPLAGPRAVLPPEPRKRALEPRATPATLASKSVSPGSLIK
TKPKKPDGNLESQHLPLNGIOTMAASVORAVTASGVPTGAVESILILQOYES
GRSKKCIQVGEFTYFNKFEDEPSGNLKNKASGSLKPVNRAKAGQVITPGDEQVGS
QOCQVPLPSPEPQVKNDECAVCHDGBELICDDCPRAFLACLSPLQETPSGLMR
CSKCIQVQNLISQPEVSRPELPALPTGPAISARCSVCGDTEVLCAHCAA
AFMRCHPPTAARPGTINLRCSKSDSPTEGTGPEAVTSGPRAPFELATKVGDDSA
SHDPLVLRHDLSESLNHSFDDILQALISMRPLAETPPFS"

ORIGIN

Query Match 10.3%; Score 159; DB 10; Length 1625;
Best Local Similarity 70.3%; Pred. No. 7.8e-21;
Matches 213; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 452 TGGAGAGAGTGTGCTCTCAGAGAGGTGCTGACCCAGCCGATCTGATGGCGCTTC 511
DB 858 TGGAGAGAGTGTGCTCTCAGAGAGGTGCTGACCCAGCCGATCTGATGGCGCTTC 917
QY 512 TTGCTGTGCGCAAGAAATGAGACGAGTGTGCTGCTGTGCGGACGCGGCGAGCTCAT 571
DB 918 CCAAGTTAACCAAGAAAGAGAGATGAGTGTGCTGCTGTGCGGACGCGGAGTGA 977
QY 572 CTGCTGTGAGCGCTGCTGCGGCGCTTCCAGCTGCGCGCTGCTGCGCGCTCGGAG 631
DB 978 CTGCTGTGAGCGCTGCTGCGGCGCTTCCAGCTGCGCGCTGCTGCGCGCTCGGAG 1037
QY 632 GATCCCAAGTGTGAGAGTGTGCTGCGCGCTGCTGCGCGCTGCTGCGCGCTCGGAG 691

Db 1038 GATCCCAAGGCGCTCTGAGATGCTCTGCTGCTCTCCAGGGGAGAGTCCACAGAACT 1097
QY 692 GCGCCGGGAGAGGACCCCCCGCCAGAGACCAACCCGTGAGAGACCCCGCTCCCCGGG 751
Db 1098 GTCCCAAGCTGAGGTGCCAGGCCCCGAGCTTACCTGAGAGACCCCGGACCAATCTCC 1157
QY 752 GCT 754
|||
Db 1158 GCT 1160

RESULT 29
AF128122 1637 bp mRNA linear ROD 29-FEB-2000
LOCUS Mus musculus autoimmune regulator (Aire) mRNA, alternatively
DEFINITION spliced product 3a, complete cds.
ACCESSION AF128122
VERSION AF128122.1 GI:7108545
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
AUTHORS 1 (bases 1 to 1637)
TITLE Expression and alternative splicing of the mouse autoimmune
regulator gene (Aire)
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1637)
Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
TITLES Direct Submission
Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW
Archer Road, Gainesville, FL 32610, USA
FEATURES
source
1..1637
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1637
/gene="Aire"
58..1287
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 3a"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAF36467.1"
/db_xref="GI:7108546"
/translation="MAGSDGMRLRLRLHRTIAVAIDSAFPLIHALADHYVPEK
QETLRLKEGCPQAFHALSLVLTDSGAILDFWILFKDYNLERSLHSLIDGFP
KQVDNLQSRKRKPLAGPRAAVLPKPPKRALEPRATPPATLASKSVSFSLK
TKPKPKDGNLESQHLPLNGIQMAVQRAVTVAAGDVPETGAVGILIQVYES
GSKKCIQVGEFTYPNKPEDPSGNLKNKARSGLKPVVRAKGAQVITPGDEQV
GOCGVPLPSLSEPOVONDECAVCHGGELICDDCPRAFHACLSPLQETPS
GLMKSCCCIQRVQONTLSPBVSRLPRLPMTDQSLQILCRDLSHARTGRSCTH
LWAPSTWACQGRRLC"

ORIGIN
Query Match 10.3%; Score 159; DB 10; Length 1637;
Best Local Similarity 70.3%; Pred. No. 7.8e-21;
Matches 213; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 572 CTGCTGTAGCGAGCTGCCCTCGGAGCTTCCACCTGAGCTGCTCCCTCCGCTCGGGA 631
|||
Db 990 CTGTTGTAGCGAGCTGTCCCCGGGCTTCCACCTGAGCTTCTCCCACTTGTGAGGA 1049
QY 632 GATCCCAAGTGGACCTGTGAGGTGCTCCAGCTGCTGAGGCAACAAGTCCAGAGGTGA 691
|||
Db 1050 GATCCCAAGTGGACCTGTGAGGTGCTCCAGCTGCTGAGGCGAGAGTCCACAGAACT 1109
QY 692 GCGCCGGGAGAGGACCCCCCGCCAGAGGACCCGCGAGACCCCGCTCCCCGGG 751
|||
Db 1110 GTCCCAAGCTGAGGTGCCAGGCCCCGAGCTTACCTGAGAGACCCCGGACCAATCTCC 1169
QY 752 GCT 754
|||
Db 1170 GCT 1172

RESULT 30
AF128125 1622 bp mRNA linear ROD 29-FEB-2000
LOCUS Mus musculus autoimmune regulator (Aire) mRNA, alternatively
DEFINITION spliced product 3d, complete cds.
ACCESSION AF128125
VERSION AF128125.1 GI:7108551
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
AUTHORS 1 (bases 1 to 1622)
TITLE Expression and alternative splicing of the mouse autoimmune
regulator gene (Aire)
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1622)
Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
TITLES Direct Submission
Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW
Archer Road, Gainesville, FL 32610, USA
FEATURES
source
1..1622
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1622
/gene="Aire"
58..1272
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 3d"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAF36470.1"
/db_xref="GI:7108552"
/translation="MAGSDGMRLRLRLHRTIAVAIDSAFPLIHALADHYVPEK
QETLRLKEGCPQAFHALSLVLTDSGAILDFWILFKDYNLERSLHSLIDGFP
KQVDNLQSRKRKPLAGPRAAVLPKPPKRALEPRATPPATLASKSVSFSLK
TKPKPKDGNLESQHLPLNGIQMAVQRAVTVAAGDVPETGAVGILIQVYES
GSKKCIQVGEFTYPNKPEDPSGNLKNKARSGLKPVVRAKGAQDEQVGOQCG
VPLPLSLSEPOVONDECAVCHGGELICDDCPRAFHACLSPLQETPSGLMRC
SCCLOGRVQONTLSPBVSRLPRLPMTDQSLQILCRDLSHARTGRSCTH
LWAPSTWACQGRRLC"

ORIGIN
Query Match 9.6%; Score 149; DB 10; Length 1622;
Best Local Similarity 78.2%; Pred. No. 6.9e-19;
Matches 179; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

[illegible]

RESULT	31			
AF128123				
LOCUS	AF128123	1634 bp	mRNA	linear
DEFINITION	Mus musculus autoimmune regulator (Aire) mRNA, alternatively spliced product 3b, complete cds.			ROD 29-FEB-2000
ACCESSION	AF128123			
VERSION	AF128123.1	GI:7108547		
KEYWORDS				
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			

REFERENCE	1 (bases 1 to 1634)
AUTHORS	Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
TITLE	Expression and alternative splicing of the mouse autoimmune regulator gene (Aire)
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1634)
AUTHORS	Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
TITLE	Direct Submission
JOURNAL	Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA
FEATURES	Location/Qualifiers
SOURCE	1..1634

```

FEATURES
Source
Location/Qualifiers
1..1634
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6"
/db_xref="taxon:10090"
/Chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1634
/gene="Aire"
58..1284
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 3b"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AF36466.1"
/db_xref="GI:7108548"
/translation="MAGDGMRLRLRLHRTETIAVAIDSAFPLILHALADHDVEDEKRF
QETLLKEKEGPQAFHALISWMLTRDGSALIDFERILFKDYNLERYSRLSILDGFPE
KVDLNGSRGRKRLPLAGRKAIVLPRTGKRLAEPPATLASGVSSPSHLK
TKPEKPDGNGLESOHLPLNGIQTMAVORAVYASGDVBTGAVGILIQOVFES
GRSKLQDGEFETYPNKPEDPSGNLKKKASGSLKRVAKGAQVITPBDKQVQ
QCCGVPPLPSLPSEFOVQNDDECAVCHDGEGLICDCCCPAAPHACISPPDQETPS
LWRSCTCQGRVQQLSQPEVSRPELPAETPDQSPLOILLCLDLSHARHTRGCTHIL
WAPSCVCCQGRGLC"

```

	Query Match	9.6%	Score 149;	DB 10;	Length 1634;
	Best Local Similarity	78.2%	Pred. No. 6.9e-19;		
	Matches	179;	Conservative	0; Mismatches	50; Indels
				0; Gaps	0;
QY	526	AGAAATGAGACGAAGTGTCCCTGTGTGTGTGGAGAGCGCGGAGACTTCATTCGCTGTGACGGCT	585		
Db	941	AGAAAGAGAGATTAGTGTCGCCGTGTGCACGACGAGAGAGTAGCTCATTCGTGTGTGACGGCT	1000		

QY	586	GCCCTCGGGACCTTACACCTGGCCGCTGATCCCTCCGCTCCGGAGATATCCCAATGGGA	645
Db	1001	GTCCCCGGACCTTTCACCTGGCTTGCTGTGCCCACTCTGCAGAGATATCCCAATGGCC	1060
QY	646	CCTGAGAGTGCTCCAGCTGCTGTGACGGCAACATGTCAAGAGTGTCAGCCCGGGCAGAGG	705
Db	1061	TCTGGAGATGCTCTGTGCTCTCCAGGGCAGATGTCAACAGAACCTGTCTCCAGGCTAGG	1120
QY	706	AGCCCCGGCCCCAGAGAGCAACCCGTGAGAACCCCGCTCCCCCGGGGCT	754
Db	1121	TGTCAGAGCCCCGGAGACTACTGTGAGAGAACCCCGAGCAATCTCCGCT	1169

RESULT	32
AX344994	
LOCUS	AX344994
DEFINITION	Sequence 65 from Patent WO0200928.
ACCESSION	AX344994
VERSION	AX344994.1 GI:18492880
KEYWORDS	
SOURCE	synthetic construct
ORGANISM	synthetic construct
	artificial sequences.

REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patient: WO 0200928-A 65 03-JAN-2002;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
SOURCE 1. .9876

```
ORIGIN
/note="chemically treated genomic DNA (Homo sapiens)"
```

Query Match	7.7%	Score 118.4	DB 6	Length 9876
Best Local Similarity	76.0%	Pred. No. 4.7e-13		
Matches 146	Conservative 0	Mismatches 46	Indels 0	Gaps 0

QY	1	AGAGAAAGTAGAGCTTCCTCAGAGCTCTTAAAGCATGGCGTTTGATGCAGAGCTTACCG	60
Db	9386	AGAGAAAGTAGAGCTTTTTTTAGGTTTTTAAAGATAGGCGTTTGATGTATTCG	9445
QY	61	CTGCTCTCAGCTGAGCCCGTGGGTGGGCGGGGCGCCCTGCTATAGCCAGAGGTCAGAG	120
Db	9446	TTGTTTTTACGTGGGTTGTCGGTGGGCGGGCGCTTTTGTATATAGTTAGGAGGTTAAG	9505
QY	121	ATTCACATGGGAATGCATGCTCATCTTTTCGTCGCCAGCATGTTTCTTAATGGGGTAGAA	180
Db	9506	ATTATTATGGGGAATGTATATTTATTTTCGTTTATAGATGTTTTTAAATGGGGTAGAA	9565
QY	181	GCAGGTGGGAG 192	
Db	9566	GTAGGTGGGAG 9577	

RESULT	33	
AX34495/c		
LOCUS		
DEFINITION	AX34495	9676 bp
ACCESSION	Sequence 66 from Patent WO0200928.	DNA
VERSION	AX34495.1	linear
	GI:18492881	PAT 01-FEB-2002

ORGANISM	synthetic construct
REFERENCE	artificial sequences.
AUTHORS	1
TITLE	Olek,A., Piepenbrock,C. and Berlin,K.
JOURNAL	Diagnosis of diseases associated with the immune system
FEATURES	Patent: WO 0200928-A 66 03-JAN-2002;
	Epigenomics AG (DE)
	Location/Qualifiers

source
1. .9876
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match
Best Local Similarity 69.2%; Pred. No. 8e-10; Length 9876;
Matches 139; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 AGAAGAGTGGAGTCTTCTCAGGCTCTTAAGAGCATGGGCTGTGCTCAGGCTGACCGG 60
Db 491 AAAAAAAAAATTAATCTTCTCAAACTTTAAAAACATAGCTTTAATCCAAATATACCGG 432

QY 61 CTGCTCTCAGTGGGCGCCGTGGGCGCGCGCCCTGCTATAGCCAGAGGTCAAG 120
Db 431 CTACTCTCAACTAAACCCGTAATAAACGAAACGCCCTACTATACCAAAAAATCAAA 372

QY 121 ATCCACTGGGAATGCCATGCTATCTTGTCCCGCATGCTTCTTAATGGGGTAGAA 180
Db 371 ATCCACTTAAAAATACCACTACTCTTGTCCCAACATTAATTTCTTAATAAAAATAAA 312

QY 181 GCAGGTCCGGAGAGACCTCC 201
Db 311 ACAATATAAAAAATACTTACC 291

RESULT 34
BV089377 581 bp DNA linear STS 15-OCT-2003
LOCUS RPAMWSE0001273 Roche Palo Alto Mus musculus STS genomic, sequence
DEFINITION
tagged site.
ACCESSION
BV089377 GI:37666856
VERSION
BV089377.1
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Usuka, J., Liao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A.,
McPherson, J.D., Foernzler, D. and Peltz, G.
Mus musculus SNPs
TITLE
Unpublished (2003)
JOURNAL
COMMENT
Contact: Jonathan Usuka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Usuka@roche.com
Primer A: No primer submitted
Primer B: No primer submitted.
Location/Qualifiers
1. .581
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="10-11206-10631-CAA01111727.1.1.11206"
/clone_lib="Roche Palo Alto"
/note="SNPs developed from as assay sequences derived from 15
different strains-of mice (as of October 1, 2003). Those
strains include A/J, A/Heu, -129/Sv, AKR/J, B10.D-H2/oshu,
BALB/cByd, BALB/cJ, C3H/HeJ, C57BL/6J, -CAST/Ei, DBA/2J,
MRL/MpJ, NZB/BinJ, NZW/Lac, SPRET/Ei..."
1. .>581

STS
Query Match
Best Local Similarity 85.6%; Score 96.6; DB 11; Length 581;
Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

ORIGIN

QY 502 TGGAGGCTCTTGTCCGTGGCCAGAGAAATGAGACAGAGTGTCCGCTGTGCGGACGCG 561
Db 274 TGGATCCCTCTTGTCCGTG-GTCAAGAGAGAGATGATGTCTCCTGTGACACGAGAG 332

QY 562 GGAAGCTCATCTGCTGTGACGCGCTCCCTCGGCGCTTCCACCTGACCTGCTGCTCCCTG 621
Db 333 GTGAGCTCATCTGTTGTGACGCGCTGTCCCGGCGCTTCCACCTGACCTGCTGCTCCCG 392

QY 622 CGCTCCGGAGATCCCCAG 640
Db 393 CTCTGACGAGATCCCCAG 411

RESULT 35
AF105002 18351 bp DNA linear ROD 26-OCT-2001
LOCUS AF105002
DEFINITION
Mus musculus autoimmune regulator (Aire) gene, complete cds.
ACCESSION
AF105002
VERSION
AF105002.1 GI:5669675
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Shi, J.D., Wang, C.Y., Marron, M.P., Ruan, Q.Q., Huang, Y.Q.,
Deter, J.C., and She, J.X.
Chromosomal localization and complete genomic sequence of the
murine autoimmune regulator gene (Aire)
Autoimmunity 31 (1), 47-53 (1999)
20059142
JOURNAL
MEDLINE
PUBMED
10593569
REFERENCE
2 (bases 1 to 18351)
Shi, J.D., Wang, C.Y., Marron, M.P., Ruan, Q.Q., Huang, Y.Q.,
Deter, J.C., and She, J.X.
Direct Submision
TITLE
Submitted (06-NOV-1998) Department of Pathology, Immunology and
Laboratory Medicine, University of Florida, 1600 SW Archer Rd.,
Room D6-15, Gainesville, FL 32610, USA
JOURNAL
COMMENT
Location/Qualifiers
1. .18351
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="10"
/feature="Aire"
/gene="Aire"
/join("<4918. .5052,5365. .5539,5769. .5925,6219. .6292,
6740. .6853,8274. .8422,8611. .8691,10373. .10488,
11577. .11676,12530. .12706,13659. .13780,13879. .13987,
16489. .16557,18058. .>18138)
/gene="Aire"
/product="autoimmune regulator"
/join(4918. .5052,5365. .5539,5769. .5925,6219. .6292,
6740. .6853,8274. .8422,8611. .8691,10373. .10488,
11577. .11676,12530. .12706,13659. .13780,13879. .13987,
16489. .16557,18058. .18138)
/gene="Aire"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAD46421.1"
/translation="MAQGDGMRLRLHRTETIAVAIDSAAPLHAAADHDVYEDKPF
DTLTLKEEGCPQAFHLLSLMLTRDGAIIIDPRRIIFKQYNLERYSLRHSIDGPF
KQVDLNSQRKGRPLAAGPAAVLPRLPKRALEPRATPATLASSVSSPSHLK
TTPPKRDPGNLESQHLPLGNGIQTAAVORAVTVASDVPRTGAVGILIQVEES
GSKKCIQVGEFYPNKFEDPSGNLKKKASGSLAKEVPAKQAVTIPGRDEQVGS
QOCQGPLPSLSEPPVONKNDCAVCHDGEELICDGPRAFLACISPLQIAPS
GLMRSCTCQGRVQONTLSQPRVSRPPLPATRPITVIGURASAEKTRGSRRLKASDA
AVTVNLAPHPAPALPRLSALCPLLSKGNBRPAPASARCVCGDGTETVLRCHKA
AAFWRCHEPFLAARPGTNLRKSCSADSTPPGTPGAIVPTSGRPAVGLAKVGDS
ASHDVLVLRHDLSESLNHSFDGILQWAIQSMRSRLAETPPSS"

CDs

ORIGIN


```

exon          /number=6
              6751..6831
              /gene="Aire"
              /number=7
intron        6832..8505
              /gene="Aire"
              /number=7
repeat_unit   7002..7159
              /gene="Aire"
              /rpt_family="SINE/B4"
repeat_unit   7027..7160
              /gene="Aire"
              /rpt_family="SINE/Alu"
repeat_unit   complement(7758..7893)
              /gene="Aire"
              /rpt_family="SINE/B2"
exon          8506..8621
              /gene="Aire"
              /number=8
intron        8622..9709
              /gene="Aire"
              /number=8
repeat_unit   9312..9458
              /gene="Aire"
              /rpt_family="SINE/Alu"
exon          9710..9809
              /gene="Aire"
              /number=9
intron        9810..10660
              /gene="Aire"
              /number=9
repeat_unit   complement(10146..10356)
              /gene="Aire"
              /rpt_family="SINE/B2"
repeat_unit   complement(10359..10551)
              /gene="Aire"
              /rpt_family="SINE/B2"
exon          10661..10837
              /gene="Aire"
              /number=10
intron        10838..11786
              /gene="Aire"
              /number=10
repeat_unit   11060..11182
              /gene="Aire"
              /rpt_family="SINE/B4"
repeat_unit   complement(11453..11562)
              /gene="Aire"
              /rpt_family="LINE/L1"
exon          11787..11908
              /gene="Aire"
              /number=11
intron        11909..12004
              /gene="Aire"
              /number=11
exon          12005..12113
              /gene="Aire"
              /number=12
intron        12114..14604
              /gene="Aire"
              /number=12
repeat_unit   complement(12262..12378)
              /gene="Aire"
              /rpt_family="SINE/Alu"
repeat_unit   complement(12277..12408)
              /gene="Aire"
              /rpt_family="LTR/MaLR"
repeat_unit   complement(12415..12431)
              /gene="Aire"
              /rpt_family="SINE/B2"
repeat_unit   complement(12432..12626)
              /gene="Aire"
              /rpt_family="SINE/B2"

```

```

repeat_unit   complement(12627..12769)
Query Match          6.3%; Score 96.6; DB 10; Length 18616;
Best Local Similarity 85.6%; Pred. No. 7.4e-09;
Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;
QY 502 TGAGGCTCTTGGCTGTCACAGAAATGAGACGAGTGTCCGTGTGTCGGACGGCG 561
    |||||
DB 8484 TGGGTCCCTCTTGCCCTG-GTCAGAGAAACGAGAGTGTCCGTGTGCCACGACGAG 8542
QY 562 GCGAGCTATCTGCTGTGACGGCTGCTCGGCGCTTCCACTGCTGCTGTGCCCTTC 621
    |||||
DB 8543 GTGAGCTATCTGTGTGTGACGGCTGTCCCGGCGCTTCCACTGCTGCTGTGCCAC 8602
QY 622 CGCTCGGAGATCCCGAC 640
    |||||
DB 8603 CTCTGAGAGATCCCGAC 8621
    |||||

RESULT 37
AF073797
LOCUS          46872 bp      DNA      linear      ROD 05-ANG-1999
DEFINITION    Mus musculus cosmid MPMGc12112287 containing the syntenic region
              of the human AIRE gene, complete sequence.
ACCESSION     AF073797
VERSION       AF073797.1 GI:4091972
KEYWORDS      HTG.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 46872)
Aalonen,U., Björnsen,P., Perheentupa,J., Horelli-Kuitunen,N.,
Lehach,H. and Yaspo,M.-L.
An autoimmune disease, APECED, caused by mutations in a novel gene
featuring two PHD-type zinc-finger domains. The Finnish-german
APECED Consortium. Autoimmune
Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy
Nat. Genet. 17 (4), 399-403 (1997)
JOURNAL       98061087
MEDLINE       9398840
PUBMED        9398840
REFERENCE     2 (bases 1 to 46872)
AUTHORS       Blechschmidt,K., Schweiger,M., Wertz,K., Poulsen,R.,
              Christensen,H.M., Rosenthal,A., Lehach,H. and Yaspo,M.L.
              The mouse Aire gene: comparative genomic sequencing, gene
              organization, and expression
              Genome Res. 9 (2), 158-166 (1999)
JOURNAL       99148139
MEDLINE       10022980
PUBMED        10022980
REFERENCE     3 (bases 1 to 46872)
AUTHORS       Blechschmidt,K., Huang-WY,C., Nordstik,G., Drescher,B.,
              Rosenthal,A. and Yaspo,M.-L.
              Direct Submission
              Submitted (22-JUN-1998) Genome Analysis, Institute of Molecular
              Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source        1..46872
              /organism="Mus musculus"
              /mol_type="genomic DNA"
              /strain="129/Ola"
              /db_xref="taxon:10090"
              /note="Cosmid MPMGc12112287 originating from cosmid
              library number 121 (RZPD-Berlin). Cloning vector Lawrlist
              7."
              complement(156..327)
              /note="Genscan, score = 14.95%, comment = Internal_exon
              172 bp frame: 2 phase: 1"
              /evidence=not experimental
              complement(156..327)
              /note="GRATL, score = 99.000%, comment = excellent"
              /evidence=not experimental
              complement(156..316)
exon

```

```
/note="Xpound exon prediction, score = 93% (0%)"
/evidence=not_experimental
complement(935..1047)
/note="Genscan, score = 11.98%, comment = Internal_exon
113 bp frame: 0 phase: 2"
/evidence=not_experimental
complement(935..1047)
/note="MZF, score = 96.5%"
/evidence=not_experimental
complement(935..1047)
/note="GRAIL, score = 100.000%, comment = excellent"
complement(1129..1208)
/evidence=not_experimental
complement(1129..1208)
/note="Genscan, score = 3.57%, comment = Internal_exon 80
bp frame: 0 phase: 2"
/evidence=not_experimental
complement(1129..1208)
/note="MZF, score = 57%"
/evidence=not_experimental
complement(1129..1208)
/evidence=not_experimental
complement(1572..1634)
/evidence=not_experimental
/rpt_family="AluSp"
1595..1669
/note="MZF, score = 91.3%"
/evidence=not_experimental
complement(1805..1885)
/evidence=not_experimental
/rpt_family="AluY/FLAM"
complement(2079..2318)
/note="Genscan, score = 4.62%, comment = Internal_exon 240
bp frame: 0 phase: 0"
/evidence=not_experimental
complement(2079..2241)
/note="MZF, score = 79.6%"
/evidence=not_experimental
complement(2079..2123)
/note="GRAIL, score = 99.000%, comment = excellent"
/evidence=not_experimental
complement(2079..2121)
/note="Xpound exon prediction, score = 84% (0%)"
/evidence=not_experimental
complement(2170..2318)
/note="GRAIL, score = 64.000%, comment = good"
/evidence=not_experimental
complement(2355..2441)
/note="Genscan, score = 5.24%, comment = Internal_exon 87
bp frame: 0 phase: 0"
/evidence=not_experimental
complement(2355..2441)
/note="MZF, score = 99.1%"
/evidence=not_experimental
complement(2355..2441)
/note="GRAIL, score = 88.000%, comment = excellent"
/evidence=not_experimental
complement(2923..3023)
/note="GRAIL, score = 70.000%, comment = good"
/evidence=not_experimental
complement(3128..3173)
/note="Genscan, score = 3.08%, comment = Internal_exon 46
bp frame: 2 phase: 1"
/evidence=not_experimental
complement(3128..3173)
/note="GRAIL, score = 94.000%, comment = excellent"
/evidence=not_experimental
complement(3128..3163)
/note="MZF, score = 75.9%"
/evidence=not_experimental
3243..3410
/note="homology = 59.50%, score = 26, counts = 28"
/rpt_type=cadem
/rpt_unit="ggaagg"
complement(3847..3975)

exon
/note="Genscan, score = 1.75%, comment = Initial_exon 129
bp frame: 0 phase: 0"
/evidence=not_experimental
complement(3847..3975)
/note="GRAIL, score = 56.000%, comment = good"
/evidence=not_experimental
4343..4367
/standard_name="gap"
/note="sequence gap of unknown size"
complement(4559..4630)
/evidence=not_experimental
/rpt_family="AluSg/X"
complement(4757..4984)
/evidence=not_experimental
/rpt_family="AluG"
complement(7733..7878)
/note="GRAIL, score = 57.000%, comment = good shadow"
/evidence=not_experimental
7794..7905
/note="GRAIL, score = 52.000%, comment = good"
/evidence=not_experimental
complement(8914..9038)
/evidence=not_experimental
/rpt_family="AluSg"
9395..9666
/note="CpG island score = 0.74, GC = 63.20%, CpGs = 25;
Region: CpG island"
/evidence=not_experimental
9555..9666
/gene="Mouse homologue of the human APECD gene (AIRE;
Acc. No. Z97990)"
/number=1
/evidence=experimental
10002..10176
/gene="Mouse homologue of the human APECD gene (AIRE;
Acc. No. Z97990)"
/number=2
/evidence=experimental
10406..10561
/gene="Mouse homologue of the human APECD gene (AIRE;
Acc. No. Z97990)"
/number=3
/evidence=experimental
10943..11017
/gene="Mouse homologue of the human APECD gene (AIRE;
Acc. No. Z97990)"
/number=4
/evidence=experimental
11465..11578
/gene="Mouse homologue of the human APECD gene (AIRE;
Acc. No. Z97990)"
/number=5
/evidence=experimental
complement(12241..12306)
/evidence=not_experimental
/rpt_family="MT1D"
complement(12539..12600)
/evidence=not_experimental
/rpt_family="RNA-Cys-TGC"
complement(12664..12819)
/evidence=not_experimental
/rpt_family="AluSg"
complement(12820..12881)
/evidence=not_experimental
/rpt_family="MT1C"
12999..13147
/gene="Mouse homologue of the human APECD gene (AIRE;
Acc. No. Z97990)"
/number=6
/evidence=experimental
```

Query Match 6.3%; Score 96.6; DB 10; Length 46872; Best Local Similarity 85.6%; Pred. No. 6,4e-09; Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

exon 13336- 13416 /gene=Mouse homologue of the human APECED gene (AIRE);

QY 502 TTGGGCGTCTTGTGCGCAAGAATAGAGACGAGTGTGCGTGTGCGGACGGGG 561

Db 15069 TTGGTCCCTCTTGCTCCT-GTCAAGAGACAGAGATGATGTGCGGTGTGACAGAGAG 15122

QY 562 GGGAGCTCATCTGCTGTGACGGCTGCGCCCTCGGGCTTCCACCTGCGCTGCTGCTCC 621

Db 15128 GTGAGCTCATCTGTTTGACGGCTGTCCCGGGCGCTTCCACTGGCTGTGCTGCCAC 15187

QY 622 CGCTCCGGGAGATCCCGAG 640

Db 15188 CTCTGACAGAGATCCCGAG 15206

RESULT 38

LOCUS AC015891 158049 bp DNA linear HTG 20-NOV-2002

DEFINITION Mus musculus chromosome 10 clone RP21-522L13 map 10. *** SEQUENCING IN PROGRESS ***, 6 unordered pieces.

ACCESSION AC015891

VERSION AC015891.16 GI:25140119

KEYWORDS HTG; HTGS PHASE1; HTGS FULLTOP; HTGS ACTIVEFIN.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 158049)

AUTHORS Birren,B., Nusbaum,C. and Lander,E.

TITLE Mus musculus chromosome 10, clone RP21-522L13

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 158049)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckert,J.R., Boguslavskiy,L., Bouhassalier,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Doneall,D., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagoes,B., Heathford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Leacock,J., Lien,C., Locke,K., MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Merrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Ralley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tsafaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 158049)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,Y., Bloom,T., Boguslavskiy,L., Bouhassalier,B., Camarata,J., Chang,J., Chazaro,B., Chespel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Goid,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagoes,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lander,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Minova,T., Miengue,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., North,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tsafaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zaitoun,J., Zembek,L., Zimmer,A. and Zody,M.

```

TITLE                               Direct Submission
JOURNAL                             Submitted (20-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT                              On Nov 20, 2002 this sequence version replaced gi:25046447. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www.beg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L761
Center clone name: 522_L_13

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
      1    58675: contig of 58675 bp in length
          1    58675: gap of 100 bp
          *    58676    58675: gap of 100 bp
          *    58775: gap of 100 bp
          *    75633: contig of 16658 bp in length
          *    75634    75633: contig of 16658 bp in length
          *    75733: gap of 100 bp
          *    75734    75733: gap of 100 bp
          *    78147: contig of 2414 bp in length
          *    78148    78147: contig of 2414 bp in length
          *    78247: gap of 100 bp
          *    78248    78247: gap of 100 bp
          *    107125: contig of 28878 bp in length
          *    107126    107125: contig of 28878 bp in length
          *    107126    107125: gap of 100 bp
          *    143742: contig of 36517 bp in length
          *    143743    143742: contig of 36517 bp in length
          *    143743    143742: gap of 100 bp
          *    143843    143743: contig of 14207 bp in length.
              Location/Qualifiers
                1..158049
                  /organism="Mus musculus"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:10090"
                  /chromosome="10"
                  /map="10"
                  /clone="RP21-522L13"
                  /clone_lib="RPCI-21 Female Mouse PAC"

ORIGIN
Query Match               6.3%   Score 96.6; DB 2; Length 158049;
Best Local Similarity    85.6%; Pred. No. 5.2e-09;
Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

OY      502 TGGGCGTCTTCCTGTGCCAGATGACAGATGAGTGCCCGTGTCCGGAGAGGC 561
DB      152857 TGGGCGTCTTCCTGTGCCAGATGACAGATGAGTGCCCGTGTCCGGAGAGGC 152915
OY      562 GAGAGCTCATGTGCTGTGACGCGCTCCCTCGGCGCTTCCACCCTGAGCTGTCCCTC 621
DB      152916 GTGAGCTCATGTGCTGTGACGCGCTCCCTCGGCGCTTCCACCCTGAGCTGTCCCTC 152975
OY      622 CGCTCGGGAAGTCCCCAG 640
DB      152976 CTGTGCAGAGATCCCCAG 152994

RESULT 39
LOCUS AC138672/c
DEFINITION AC138672 190019 bp DNA linear HTG 22-MAR-2003
SEQUENCE, 13 unordered pieces.
AC138672
VERSION AC138672.3 GI:29150492
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```


REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Birren,B., Nuebaum,C. and Lander,E.
TITLE	Mus musculus chromosome 10, clone RP23-411J14
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 190019)
AUTHORS	Birren,B., Nuebaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barua,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,U., Chang,J., Chazaro,B., Chepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Haez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,T., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,S., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuppach,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmerman,A. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (15-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	3 (bases 1 to 190019)
AUTHORS	Birren,B., Nuebaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Aracich,H.M., Barua,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,U., Chang,J., Chepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Doolley,K., Dorris,L., Erickson,J., Fato,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Haez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C., MacDonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupa,A., Ramsamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuppach,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmerman,A. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (22-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Mar 22, 2003 this sequence version replaced gi:28394987. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center -----	
Center: Whitehead Institute/ MIT Center for Genome Research	
Center code: WIBR	
Web site: http://www-seq.wi.mit.edu	
Contact: sequence_submissions@genome.wi.mit.edu	
----- Project Information -----	
Center project name: I289928	
Center clone name: 411_J_14	
----- Summary Statistics -----	
Sequencing vector: Plasmid; n/a; 100% of reads	
Chemistry: Dye-terminator Big Dye; 100% of reads	
Assembly program: Phrap; version 0.960731	
Consensus quality: 186901 bases at least Q40	
Consensus quality: 187957 bases at least Q30	
Consensus quality: 188318 bases at least Q20	
Insert size: 188819; sum-of-contigs	
Quality coverage: 7.6 in Q20 Bases; sum-of-contigs	

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 11865: contig of 11865 bp in length
* 11866 11865: gap of 100 bp
* 11966 15071: contig of 3106 bp in length
* 15072 15171: gap of 100 bp
* 15172 17398: contig of 2227 bp in length
* 17399 17498: gap of 100 bp
* 17499 19092: contig of 1594 bp in length
* 19093 19192: gap of 100 bp
* 19193 22836: contig of 3644 bp in length
* 22837 22936: gap of 100 bp
* 22937 31109: contig of 8173 bp in length
* 31110 31209: gap of 100 bp
* 31210 45955: contig of 14746 bp in length
* 45956 46056: gap of 100 bp
* 46056 70470: contig of 24415 bp in length
* 70471 70570: gap of 100 bp
* 70571 97501: contig of 26931 bp in length
* 97502 97601: gap of 100 bp
* 97602 117865: contig of 20264 bp in length
* 117866 117965: gap of 100 bp
* 117966 140668: contig of 22703 bp in length
* 140669 140768: gap of 100 bp
* 140769 167104: contig of 26336 bp in length
* 167105 167205: gap of 100 bp
* 167205 190019: contig of 22815 bp in length.
*
FEATURES
source
1..190019
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosome="10"
    /map="10"
    /clone="RP23-411J14"
    /clone_1fb="RPC1-23 Female Mouse BAC"
1..11865
    /note="assembly_fragment"
    /clone_end:SP6
    /vector_side:left"
11966..15071
    /note="assembly_fragment"
15172..17398
    /note="assembly_fragment"
17499..19092
    /note="assembly_fragment"
19193..22836
    /note="assembly_fragment"
22937..31109
    /note="assembly_fragment"
31210..45955
    /note="assembly_fragment"
46056..70470
    /note="assembly_fragment"
70571..97501
    /note="assembly_fragment"
97602..117865
    /note="assembly_fragment"
117966..140668
    /note="assembly_fragment"
140769..167104
    /note="assembly_fragment"
167205..190019
    /note="assembly_fragment"
    /clone_end:T7
    /vector_side:right"

```

```

ORIGIN
Query Match      6.3%; Score 96.6; DB 2; Length 190019;
Best Local Similarity 85.6%; Pred. No. 5.1e-09;
Matches 119; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

QY 502 TGGGGGCTCTTGGCCCTGCGCAGAGATGAGACGAGTGTCCGCTGTGTCGGGAGGGG 561
DB 59846 TGGGTCTCTTGTGCTG-GTGAGAACAGAGATAGTGTGCGGTGTCACGACGAGG 59788
QY 562 GGGAGCTCATCTGTGTGAAGGCTGCGCTTCCACCTGAGCTGTGCTCCCTC 621
DB 59787 GTGAGCTCATCTGTGTGAAGGCTGCGCTTCCACCTGAGCTGTGCTCCCTC 59728
QY 622 CGCTCGGAGATCCCGAG 640
DB 59727 CTCTGAGAGATCCCGAG 59709

RESULT 40
ACI08592/c 217688 bp DNA 1linear HTG 15-NOV-2002
DEFINITION Rattus norvegicus clone CH230-144C19, *** SEQUENCING IN PROGRESS
ACCESSION ACI08592
VERSION ACI08592.5 GI:25006698
KEYWORDS HTG, HTGS_PHASE1, HTGS_DRAFT, HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 217688)
Muzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsdorfs, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Balwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devalla, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Din, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egal, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kows, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louised, H., Lozano, R. J., Lu, X., Ma, T.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawlin, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Moran, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwachukwu, O., Okunolu, G., Olarinmaga, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L.,
Puzo, M., Quiroz, J., Rachlin, R., Reeves, K., Regier, M. A., Reich, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. U.,
Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shan, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, D.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Tjorvis, Z., Umant, K.,
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Woodman, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 217688)
Worley, K. C.
Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 217688)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23270241.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPMS
Center clone name: CH230-144C19
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 177651 bases at least Q40
Consensus quality: 182606 bases at least Q30
Consensus quality: 185802 bases at least Q20
Estimated insert size: 178707; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 161680: contig of 161680 bp in length
161681 161780: gap of unknown length
161781 170889: contig of 9109 bp in length
170890 170989: gap of unknown length
170990 177789: contig of 6800 bp in length
177790 177889: gap of unknown length
177890 189681: contig of 11792 bp in length
189682 189781: gap of unknown length
189782 211300: contig of 21519 bp in length
211301 212789: gap of unknown length
212789: contig of 1389 bp in length

```



```

Qy 448 GTTTGGGAAGAGGAGGCTCTGACAGAGGCTGACACCCAGCCAGTCTGCATGGAGC 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17619 KRAKKRYKCGRRRRSSYMKCCKMMMSYCMYSCTMYYSKSTYKSSCTYRGGYWG 17560
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 508 TCTCTTCCGCTGTCAGAGATAGAGAGAGGTCGCGGTCGAGGAGGAGGAGC 567
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17559 SKTCTYAGGKSNMYCMMRSSKSSSSWMSMABSSWCMGAGATRRSKSAGAGRRSS 17500
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 568 TCATCTGCTGTAGAGGCTGCTCGGAGCCTTCCACCTGAGCTCTGCTCCCTCCGCTCC 627
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17499 KGRKSTGM--KRAKSKKTSYGSSTGRSMKKKGYSKYRSRGMKKKCTCYMMYYKRYKT 17442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 628 GGGAGATCCCACTGAGGAGCTGAGGCTCTCCAGCTGCTGCGAGCAACAGTCCAGAGG 687
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17441 SMCMMWYKMSWGYKRYKRCCKMKKGGCTGYGMSKSKSGYKSMRGMSSYSTGCMWSWG 17382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 688 TGCAGCCCGGAGAGAGAGCCCGGAGCCAGCCGCTGAGAGACCCGCTCCGCCC 747
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17381 YSMKCMKTKMSYKKRRSRMGSMSKMGAYAGRCYSSSMSTRKRKSKCYKSKYKKG 17322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 748 CCGGAGCTTAGTGGCGGAGAGAGAGTCACTGAGGAAACCCCTAGCCGCGCA 807
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17321 RGRMKMGWGMKRGSKTYMSMKMKRSMSKCYSTKSYSGRRSKGMRSTKAKSSMR 17262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 808 -TGACACGACCTTGTCTCAACAAGACCTGCGGCTCCGCTCTGAGCCCGGCTGCCA 866
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17261 MAGSKCTYSSSYNNRRNNRMGKTGCTNNYRRSRMMNMAAAGCTTCCGANTMGGG 17202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 867 GGTCTGAGACTCCCTCGGCTGACACCCCTACTGCTGCTGAGGCTGACGAGAC 926
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17201 GGAAGAGGCGGASASACSKYKMSKSKSCYRSSTRRRCKMSKTRRSRSGMSMTGRSG 17142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 927 CTGCTCTCTGTGTGCGGCTTGGGAGTGTGCGAGATGTAACGACGCTGCTGCTGCTACT 986
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17141 GKTSYBAKSSGRCCYCMWGRKGRKCMKSKYRRKRMWTGCMKCYSMWAMY 17082
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 987 CACTGCGCGGCTGCTTCCACTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1046
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17081 CRSMCMCMCKSCCGCYSMGMSYSYSGKYSWKMKSMMMSYSYKSS--TSKMRSSKRG 17024
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1047 ACAGGCTGCTGCTGACAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17023 MGTGATYKGGRRYKGGGSKYKGGGSMWGGKSGKSTGSSAGKSS---SYKMSCCARYK 15968
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1107 CTGCGCCCGGAGCCCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1166
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16967 SMSYKCYRMMWCMYYCYCMWGMSSCWSMSMRKSKSGSAKGMVMDKGSVTDKSD 16908
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1167 GAGCCCGCTCTGACAGGATGACCTGAGTCCCTTC 1203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16907 KMSRSBSKVKSMMSRGRKYYCTDMSKMYTTCG 16871
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 47
BX649390 184594 bp DNA linear VRT 11-FEB-2004
LOCUS Zebrafish DNA sequence from clone DKEY-261L2 in linkage group 2,
DEFINITION complete sequence.
ACCESSION BX649390
VERSION BX649390.12 GI:42538836
KEYWORDS HTG
SOURCE Danto rerio (zebrafish)
ORGANISM Danto rerio

REFERENCE
AUTHORS Johnson, C.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Feb 11, 2004 this sequence version replaced gi:4251008.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zf1sh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep. Clone-derived zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/projects/D_rerio/fishmark.shtml DKEY-261L2 is from a zebrafish BAC library VECTOR: pindigBAC-5.

FEATURES

source
1..184594
/organism="Danto rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-261L2"
/clone_1ib="Dantokey"

ORIGIN

Query Match 4.5%; Score 70; DB 5; Length 184594;
Best Local Similarity 74.6%; Pred. No. 0.00077;
Matches 88; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```

Qy 523 AGAAGATAGAGAGAGTGTGCTGCGGAGCGCGGAGAGTCTGCTGAGACG 582
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78081 AGCACAATATATGATGTGACGATGTGTAAGACGCGCGAGCATCTGCTGATG 78022
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 583 GCTGCGCTGCGGCTTCCACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 640
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78021 GGTGTCTGTGCTGCTTCACTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 77964
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 48
AK127046 4966 bp mRNA linear PRI 19-FEB-2004
LOCUS Homo sapiens cDNA FL45103 f18, clone BRAMH3032571, moderately
DEFINITION similar to Chromodomain helicase-DNA-binding protein 4.
ACCESSION AK127046
VERSION AK127046.1 GI:34533779
KEYWORDS oligo capping; f18 (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,

Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsuna, M., Murakawa, K., Kanenori, K., Takahashi-Fuji, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahashi, K., Masuh, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 4966)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, Flu Project (HRI Team) ; 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: Reverse Proteomics Research Institute, HRI and RAB.

FEATURES
source
Location/Qualifiers
1..4966
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BRAMH3032571"
/issue_type="brain"
/clone_lib="BRAMH3"
/note="cloning vector: pME18SFL3"
1..3678
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC86802.1"
/db_xref="GI:34533780"
/translation="MRGPVGTSEELPRLFAEMENEDENSEEEDGGLAFADDFFPYEP
VLPKKKKRKKLKENCKGKRRKKSGNDELSENEDLEKSESSEGSYSPMKKKKK
LKDCKKAKRKKKDDDEDNDGCLKEPKSGGOLMAEWGLDDVDVLFSEBYHTLTN
YAFSGOPLRLAKKPKTPMSKMTVLGAKREBSANNPFKSSAAAAAVALVAVE
TVTISPLAVSPQVPQVPIRAKATKESGQVARKIKSGDGGKGGKGTAGAKF
RFGISNNRKKSSSEDEBESDPSAISHASVRSBSALGKSKRRKKRID
GQVETDHDQYCEVCOQGEIILCDTCEPAVHLVCLDPELEKAPBGKSCPCKEGI
OWEKDDDEBEEGCEEEDDMFCRCVCKOGGELLCCDAPSSYHLCNLPLEI
PGEWMLCPRCCTCPPLKGVORLIHMRTEPPAPFMVGLGPVPELPPKPLEGIP
REFPYKALSYMHCNWKLOELIYHTMYRYQKNDMPPEPPDYSGBEDGSE
KKNKDPYLAKKEERFYRGIRPEMMHRIILNHSFDKGDVHTLTKMDLPYDQCTW
ELDDIDIPYDNLKQAVWGHRELMLGEDTGLPRLILKQKRLDKQEKRPDPIVDP
TVKFDQMPYIDSTGTLHPYQLEGLNMLRFSWAQGTDTIADEMGLGTVQTVFLY
SLYKESGSKPYLVASPLSTIINMREPEMMAADPFVVTYTGDKESRSYIRNEBSFE
DNAIRSGKVPFMKKEVOIKFHYLTSVELITIDOLIGSIEMACLVDEARHLKNO
SKPRVLNSYKIDYKLLLTGTPLONNLELFIHLNFIPIPERNNLEGLPEPRADISKE
DOIKLHDLGPHMLRLKADVFKNMPATTELIVRVLSOMQKTYKFLITRFBALN
SKGGNQVSLINIMDLKCCNHPYLPVAALVAVLPNGSYDSSLVSSSGKLMLQ
KMLKRLRDEGHRVLI.FSQMTKMLDLLEDFLEBYGYKERIDGIGTGLRQEAIDRFNA
PQAQOCPFLSTRAGGLINTLADTVLIYDSDPNPHNDIOAFSAHRI.GOTKWKGSW
ATMWRKTPGHRSGSPMKSRPQPPWIMWRVIRSTRAGPARREBSGQRRRRPRRP
RSCRRRCRCLFLRRRSWTSMWS"

CDS
Query Match 4.2%; Score 65; DB 9; Length 4966;
Best Local Similarity 62.7%; Pred. No. 0.013;
Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTGCGGTGTGCGGACGGCGGGAGCTCATCTGCTGACGGCTGCCCTC 591
DB 1028 AGGATTACTGTGAGGTGTGCGACAGGCTGGGGAGATCATCTGTGTGACACCTGCCCA 1087
QY 592 GGGCCTTCCACCGGCGGTGCTGCCCTCCGCTCCGGGAGATCCCAAGTGGGACCTCGGA 651
DB 1088 GGGCTTCAACCTTCGTATGCTTGAGACCCAGAGCTGGAAGAGGCTCCGAGGGCAAGTTGA 1147
QY 652 GGTGCTCAGCTGCTGTCAGGCAACAGTCCAGGAGGTGACG 92

Db 1148 GCTGCCCCCATCTGTGAGAAGAGGGGATCCATCGGAGCGG 1188
RESULT 49
AF425231
LOCUS
DEFINITION
Homo sapiens chromodomain helicase DNA binding protein 5 (CHD5)
mRNA, complete cds.
AF425231.1 GI:19773959
ACCESSION
AF425231
VERSION
AF425231.1
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Thompson, P.M., Gotch, T., White, P.S. and Brodeur, G.M.
TITLE
CHD5, a New Member of the Chromodomain Gene Family, is
Preferentially Expressed in the Nervous System
JOURNAL
Unpublished
2 (bases 1 to 9646)
Thompson, P.M., Gotch, T., White, P.S. and Brodeur, G.M.
Direct Submission
Submitted (27-SEP-2001) Oncology, Children's Hospital of
Philadelphia, 3615 Civic Center Blvd., Philadelphia, PA 19104-4318,
USA

FEATURES
source
Location/Qualifiers
1..9646
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="1p36.3"
1..9646
/gene="CHD5"
101..5965
/gene="CHD5"
/codon_start=1
/product="chromodomain helicase DNA binding protein 5"
/protein_id="AAL98962.1"
/db_xref="GI:19773960"
/translation="MRGPVGTSEELPRLFAEMENEDENSEEEDGGLAFADDFFPYEP
VLPKKKKRKKLKENCKGKRRKKSGNDELSENEDLEKSESSEGSYSPMKKKKK
LKDCKKAKRKKKDDDEDNDGCLKEPKSGGOLMAEWGLDDVDVLFSEBYHTLTN
YAFSGOPLRLAKKPKTPMSKMTVLGAKREBSANNPFKSSAAAAAVALVAVE
TVTISPLAVSPQVPQVPIRAKATKESGQVARKIKSGDGGKGGKGTAGAKF
RFGISNNRKKSSSEDEBESDPSAISHASVRSBSALGKSKRRKKRID
GQVETDHDQYCEVCOQGEIILCDTCEPAVHLVCLDPELEKAPBGKSCPCKEGI
OWEKDDDEBEEGCEEEDDMFCRCVCKOGGELLCCDAPSSYHLCNLPLEI
PGEWMLCPRCCTCPPLKGVORLIHMRTEPPAPFMVGLGPVPELPPKPLEGIP
REFPYKALSYMHCNWKLOELIYHTMYRYQKNDMPPEPPDYSGBEDGSE
KKNKDPYLAKKEERFYRGIRPEMMHRIILNHSFDKGDVHTLTKMDLPYDQCTW
ELDDIDIPYDNLKQAVWGHRELMLGEDTGLPRLILKQKRLDKQEKRPDPIVDP
TVKFDQMPYIDSTGTLHPYQLEGLNMLRFSWAQGTDTIADEMGLGTVQTVFLY
SLYKESGSKPYLVASPLSTIINMREPEMMAADPFVVTYTGDKESRSYIRNEBSFE
DNAIRSGKVPFMKKEVOIKFHYLTSVELITIDOLIGSIEMACLVDEARHLKNO
SKPRVLNSYKIDYKLLLTGTPLONNLELFIHLNFIPIPERNNLEGLPEPRADISKE
DOIKLHDLGPHMLRLKADVFKNMPATTELIVRVLSOMQKTYKFLITRFBALN
SKGGNQVSLINIMDLKCCNHPYLPVAALVAVLPNGSYDSSLVSSSGKLMLQ
KMLKRLRDEGHRVLI.FSQMTKMLDLLEDFLEBYGYKERIDGIGTGLRQEAIDRFNA
PQAQOCPFLSTRAGGLINTLADTVLIYDSDPNPHNDIOAFSAHRI.GOTKWKGSW
ATMWRKTPGHRSGSPMKSRPQPPWIMWRVIRSTRAGPARREBSGQRRRRPRRP
RSCRRRCRCLFLRRRSWTSMWS"

ORIGIN

MAGNKPANALVHKVLEKLBELLSDMKADVTBPLPATIARIPIVAVRLQMSERNLSRLA
NEAPEPPEQOVAQOQ"

Query Match 4.1%; Score 63.2; DB 10; Length 6438;
Best Local Similarity 67.4%; Pred. No. 0.028;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GAGCAGTGTGCGCTGTGTGCGGACGGCGGAGGAGTCTGTGTGTGAGCGGCTGCCCTCG 592
DB 1503 GGAATTTCTGTGCGGTGTGCAAGAGCGCGGAGGCTCTGTGTGTGACACATGCCCTTC 1562
QY 593 GGCCTTCCAGCTGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 652
DB 1563 TTCTTACCACTTCACATGCTGCTGAACCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1622
QY 653 GTGCTCCAGCTG 664
DB 1623 CTGTCCCGCTG 1634

RESULT 51
AC122760 67573 bp DNA linear HTG 15-JUN-2002
LOCUS Mus musculus clone RP24-489C24, LOW-PASS SEQUENCE SAMPLING.
AC122760
AC122760.2 GI:21427758
VERSION HTG, HTGS_PHASE0.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Birren, B., Linton, L., Nusbbaum, C. and Lander, E.
JOURNAL Mus musculus, clone RP24-489C24
REFERENCE 1 (bases 1 to 67573)
AUTHORS Unpublished
2 (bases 1 to 67573)

REFERENCE
AUTHORS Birren, B., Linton, L., Nusbbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Bouhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Deatrelano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
Grand-pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labocque, K.,
Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,
Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicoll, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE Direct Submision
JOURNAL Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome
RESEARCH Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 67573)
AUTHORS

TITLE
JOURNAL

COMMENT

Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,
Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicoll, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submision
Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 15, 2002 this sequence version replaced gi:21206372.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26032
Center clone name: 489_C_24

NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
712 811: contig of 711 bp in length
812 811: gap of 100 bp
812 1523: contig of 712 bp in length
1524 1623: gap of 100 bp
1624 2341: contig of 718 bp in length
2342 2441: gap of 100 bp
2442 3148: contig of 707 bp in length
3149 3248: gap of 100 bp
3249 3962: contig of 714 bp in length
3963 4062: gap of 100 bp
4063 4778: contig of 716 bp in length
4779 4878: gap of 100 bp
4879 5593: contig of 715 bp in length
5594 5693: gap of 100 bp
5694 6406: contig of 713 bp in length
6407 6506: gap of 100 bp
6507 7225: contig of 719 bp in length
7226 7325: gap of 100 bp
7326 8023: contig of 698 bp in length
8024 8123: gap of 100 bp
8124 8820: contig of 697 bp in length
8821 8920: gap of 100 bp
8921 9619: contig of 699 bp in length
9620 9719: gap of 100 bp
9719 10417: contig of 698 bp in length
10418 10517: gap of 100 bp
10518 11224: contig of 707 bp in length
11225 11324: gap of 100 bp
11325 12037: contig of 713 bp in length
12038 12137: gap of 100 bp
12138 12837: contig of 700 bp in length
12838 12937: gap of 100 bp
12938 13661: contig of 724 bp in length
13662 13761: gap of 100 bp

```

* 13762 14487: contig of 726 bp in length
* 14488 14587: gap of 100 bp
* 14588 15291: contig of 704 bp in length
* 15292 15391: gap of 100 bp
* 15392 16079: contig of 688 bp in length
* 16080 16179: gap of 100 bp
* 16180 16877: contig of 698 bp in length
* 16878 16977: gap of 100 bp
* 16978 17655: contig of 678 bp in length
* 17656 17755: gap of 100 bp
* 17756 18470: contig of 715 bp in length
* 18471 18570: gap of 100 bp
* 18571 19236: contig of 666 bp in length
* 19237 19336: gap of 100 bp
* 19337 20042: contig of 706 bp in length
* 20043 20142: gap of 100 bp
* 20143 20842: contig of 700 bp in length
* 20843 20942: gap of 100 bp
* 20943 21660: contig of 718 bp in length
* 21661 21760: gap of 100 bp
* 21761 22475: contig of 715 bp in length
* 22476 22575: gap of 100 bp
* 22576 23281: contig of 706 bp in length
* 23282 23381: gap of 100 bp
* 23382 24097: contig of 716 bp in length
* 24098 24197: gap of 100 bp
* 24198 24898: contig of 701 bp in length
* 24899 24998: gap of 100 bp
* 24999 25689: contig of 691 bp in length
* 25690 25789: gap of 100 bp
* 25790 26489: contig of 700 bp in length
* 26490 26589: gap of 100 bp
* 26590 27288: contig of 699 bp in length
* 27289 27388: gap of 100 bp
* 27389 28077: contig of 689 bp in length
* 28078 28177: gap of 100 bp
* 28178 28881: contig of 704 bp in length
* 28882 28981: gap of 100 bp
* 28982 29692: contig of 711 bp in length
* 29693 29792: gap of 100 bp
* 29793 30513: contig of 721 bp in length
* 30514 30613: gap of 100 bp
* 30614 31334: contig of 721 bp in length
* 31335 31434: gap of 100 bp
* 31435 32135: contig of 701 bp in length
* 32136 32235: gap of 100 bp
* 32236 32931: contig of 696 bp in length
* 32932 33031: gap of 100 bp
* 33032 33742: contig of 711 bp in length
* 33743 33842: gap of 100 bp
* 33843 34542: contig of 700 bp in length
* 34543 34642: gap of 100 bp
* 34643 35354: contig of 712 bp in length
* 35355 35454: gap of 100 bp
* 35455 36140: contig of 686 bp in length
* 36141 36240: gap of 100 bp
* 36241 36953: contig of 713 bp in length
* 36954 37053: gap of 100 bp
* 37054 37747: contig of 694 bp in length
* 37748 37847: gap of 100 bp
* 37848 38564: contig of 717 bp in length
* 38565 39372: gap of 100 bp
* 39373 39472: gap of 100 bp
* 39473 40178: contig of 706 bp in length
* 40179 40278: gap of 100 bp
* 40279 40987: contig of 709 bp in length
* 40988 41087: gap of 100 bp
* 41088 41809: contig of 722 bp in length
* 41810 41909: gap of 100 bp
* 41910 42601: contig of 692 bp in length
* 42602 42701: gap of 100 bp
* 42702 43395: contig of 694 bp in length

```

```

Query Match
Best Local Similarity 67.4%; Score 63.2; DB 2; Length 67573;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGATGTCGTCGTGTCGGAGCGCGGAGCATCTGTCGTGACGCTCCCTCG 592
Db 12393 GGAATTCGTTCGGCTGTGCAAGAGCGCGGAGCTCCGTGTCGTGACATGCTTC 12452
QY 593 GGCCTTCACCTGACCTGCTCTCCCTCCCTCGCGGAGATCCCAATGCGGACCTGAG 652
Db 12453 TTCCTACCAATCCACTGCTGCAACCCCGCTGCGAGATCCCAAGCGGAATGCT 12512
QY 653 GTCCTCCAGCTG 664
Db 12513 CTGTCCCGCTG 12524

RESULT 52
AC134529
LOCUS
DEFINITION
AC134529 173909 bp DNA linear HTGS 05-MAR-2003
MUS musculus chromosome UNK clone RP24-503A2, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
AC134529
AC134529.3 GI:28850175
HTGS HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS
MUS musculus (house mouse)
SOURCE
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 173909)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 173909)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 173909)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (05-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Mar 5, 2003 this sequence version replaced gi:28626883.

REFERENCE
AUTHORS
TITLE
JOURNAL
JOURNAL
COMMENT

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BB0503A02
----- Summary Statistics -----
Sequencing vector: pLJ3; 0%
Sequencing vector: pLJ3; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175465 bases at least Q40
Consensus quality: 176221 bases at least Q30
Consensus quality: 176527 bases at least Q20
NOTE: This is a 'working draft' sequence. It currently
consists of 5 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

MOLGAU, M., MOLLER, N., MOLLER, N.

[illegible]

Nankevis, C., Nea, D., Newton, N., Nguyen, N., Norris, S.,
 Nwokwelen, C., Okunnu, G., Olanrewaju, A., Pal, S., Parks, K.,
 Paeternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Ploster, F., Poldinger, A., Popovic, D., Primus, E., Pu, L.,
 Plazo, M., Quirós, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Kelly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
 Sanders, M., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
 Shetty, J., Shvatsbeyn, A., Sison, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sores, J.,
 Steinle, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K.,
 Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, J.,
 Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Unpublished
 2 (bases 1 to 248835)
 Morley, K.C.
 Direct Submission
 Submitted (13-OCT-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 248835)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 10, 2003 this sequence version replaced g1:231100986.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: G1R1
 Center clone name: CH230-172E6
 ----- Summary Statistics
 Assembly program: Atlas 3.0
 Consensus quality: 212993 bases at least Q40
 Consensus quality: 215360 bases at least Q30
 Consensus quality: 217739 bases at least Q20
 Estimated insert size: 225359; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 23011: contig of 23011 bp in length

DAFTOMLVIRDRGSEKEFKAYVSLFEMHLCEPDADGETRADGPREGCSROHYLT
RIGWMSLIRKYOBERHNVGRMSMPELAVENKMSQSGSPPTPTSTGTGDTOPN
TPAPVPADGDKIENSJLKEESIEGEKREVSIVAPETVIECTQAPAPSEKXVVE
PEGEKREKAVKERTERPEBTEPGADVKEVSKAIDLTPIVDEKKEKEEB
KKEVMDONGETPDLNDEKOKNIKORFMFNALNDEGTELHLMQHEERAAVTTKTY
EIMHRHDLWMLAGIINNGYAMODIONPRAVALNEPKGMNNGNFLEIKNFKLAR
RKLEALVIEBOLRRAYLMSBSPMSALNTRPAVEVCIAESHOHLSKESMAG
MKPANAVALHUKOLEBELLSDMKADVTPLPATIATLPVAVRLQMSERNILSRLANRA
PEPTPOVAQOO

ORIGIN

Query Match 3.9%; Score 60; DB 9; Length 6417;
Best Local Similarity 61.5%; Pred. No. 0.12;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGTGCGGACGCGGGAGACTCATCTGTGTGAAGGCTGCCCTC
1198 AGGACTATTGCGAGGTGTGCGCAGCAAGCGGTGAGATCATCTGTGTGATTAAGTCCCTC
DB 1198 AGGACTATTGCGAGGTGTGCGCAGCAAGCGGTGAGATCATCTGTGTGATTAAGTCCCTC
QY 592 GGGCCCTTCGACCTGGCTGCTGCTCCCTCCCTCCGCGGAGATCCCAAGTGGACCTGGA
1258 GGTGCTTACCAATGCTGCTGCTGATCCGACATGAGGAAGGCTCCCGAGGCGCAAGTGA
DB 1258 GGTGCTTACCAATGCTGCTGCTGATCCGACATGAGGAAGGCTCCCGAGGCGCAAGTGA
QY 652 GGTGCTCAGCTGCTGCTGAGCAAGCAAGTCCAGAGG 687
1318 GCTGCCCACTGCGAGGAAGCAAGCATCAAGTGG 1353
DB 1318 GCTGCCCACTGCGAGGAAGCAAGCATCAAGTGG 1353

RESULT 65
AR338834 6475 bp DNA linear PAT 17-AUG-2003
LOCUS AR338834 Sequence 325 from patent US 6569662.
DEFINITION AR338834
ACCESSION AR338834
VERSION AR338834.1 GI:33725691
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6475)
AUTHORS Tang,Y.T., Zhou,P. and Drmanac,R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 6569662-A 325 27-MAY-2003;
FEATURES
source 1..6475
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 3.9%; Score 60; DB 6; Length 6475;
Best Local Similarity 61.5%; Pred. No. 0.12;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGTGCGGACGCGGGAGACTCATCTGTGTGAAGGCTGCCCTC
1327 AGGACTATTGCGAGGTGTGCGCAGCAAGCGGTGAGATCATCTGTGTGATTAAGTCCCTC
DB 1327 AGGACTATTGCGAGGTGTGCGCAGCAAGCGGTGAGATCATCTGTGTGATTAAGTCCCTC
QY 592 GGGCCCTTCGACCTGGCTGCTGCTCCCTCCCTCCGCGGAGATCCCAAGTGGACCTGGA
1387 GGTGCTTACCAATGCTGCTGCTGATCCGACATGAGGAAGGCTCCCGAGGCGCAAGTGA
DB 1387 GGTGCTTACCAATGCTGCTGCTGATCCGACATGAGGAAGGCTCCCGAGGCGCAAGTGA
QY 652 GGTGCTCAGCTGCTGCTGAGCAAGCAAGTCCAGAGG 687
1447 GCTGCCCACTGCGAGGAAGCAAGCATCAAGTGG 1482
DB 1447 GCTGCCCACTGCGAGGAAGCAAGCATCAAGTGG 1482

RESULT 66
BC038596 6584 bp mRNA linear PRI 07-OCT-2003
LOCUS BC038596
DEFINITION Homo sapiens chromodomain helicase DNA binding protein 4, mRNA
(cDNA clone MGC:46187 IMAGE:5528023), complete cds.
ACCESSION BC038596
VERSION BC038596.1 GI:24047225
KEYWORDS MGC.

SOURCE

ORGANISM

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 6584)
Struhsberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schert,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulik,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kerteman,M., Madan,A., Young,A.C., Shevchenko,Y.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Boutard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywicki,M.I., Skalek,U., Smalins,D.E.,
Scherer,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22386257
MEDLINE
JOURNAL
TITLE
PUBMED
2 (bases 1 to 6584)
Struhsberg,R.
REFERENCE
Submitted (15-OCT-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@ncl.nih.gov
Ahter,N., Aylee,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dieterich,N.L., Granite,S., Guan,X., Gupta,U., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Larc,P., Legaspi,R.,
Maduro,O.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,U., Pearson,R., Stantripp,S., Thomas,P.J., Touchman,J.W.,
Tsuang,C., Vogt,J.L., Walker,M.A., Wechetby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
Series: IRAX Plate: 78 Row: 0 Column: 18
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4557452.
Location/Qualifiers

FEATURES

source

1..6584
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:46187 IMAGE:5528023"
/cissue_type="Skin, melanotic melanoma."
/clone_id="NIH_MGC_72"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..6584
/gene="CHD4"
/note="synonym: M12-BETA"

gene


```

FEATURES
source
* 130207 132592: config of 2386 bp in length.
location/Qualifiers
1..132592
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.21-p36.33"
/clone="RP5-889P23"
/clone_1fb="RPIC1-5"
1..3331
/note="assembly_fragment:00799"
misc_feature
3432..90998
/note="assembly_fragment:02114"
misc_feature
91099..130106
/note="assembly_fragment:03033.0"
misc_feature
130207..132592
/note="assembly_fragment:03380"

```

Query Match	Similarity	Score	DB	Length
Best Local	65.9%	Pred. 0.072		
Matches	87	Conservative	0	Mismatches 45; Indels 0; Gaps 0;
QY	533	GAACAGAGTGCCTGTTGTGGGAGCGCGGGAGCTATCTGCTGTGACGCGCTGGCCCTGG	592	
DB	33964	GAAGTTCTGCCCCGCTGTGCAGAGAGCGGGGGAGCTCTTGTCTGCACGCTGGCCCTTC	34023	
QY	593	GAGCTTCACCTGGCCCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGACTTGGAG	652	
DB	34024	CTCTTACACCTGTGATTGTGCTCAACCGCGCGCTGCCGAGATCTCCAAACGTGAATGGCT	34083	
QY	653	GTGCTCCAGCTG	664	
DB	34084	CTGCCCGGCTG	34095	

RESULT 68	
LOCUS	HS120G22
DEFINITION	HS120G22 166518 bp DNA linear PRT 21-JUL-2000
ACCESSION	Human DNA sequence from clone Rpi-120G22 on chromosome 1p36.21-36.33, complete sequence.
VERSION	AL031847
KEYWORDS	AL031847.17 GI:9369286
SOURCE	HTG.
ORGANISM	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 166518) Wray, P.
REFERENCE	Direct Submision
AUTHORS	Submitted (20-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
JOURNAL	On Jul 22, 2000 this sequence version replaced gi:9368784.
COMMENT	

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormed This sequence

was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chr1>
RP1-120G32 is from the library RPCL-1 constructed at the Roswell Park Cancer Institute by the group of Pter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RP1-120G32. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true right end of clone RP1-120G32 is at 166518 in this sequence. The true left end of clone R55-889P23 is at 96508 in this sequence. The true right end of clone RP1-20208 is at 100 in this sequence.

URES	Location/Qualifiers
source	1. 166518 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="1" /map="p36.21-36.33" /clone="RP1-120G22" /clone_id="RPCT-1" 8814. .9131
misc_feature	/note="match: GSS: Em:A0716455" 8816. .9229
misc_feature	/note="match: GSS: Em:A0239786" 10012. .10541
misc_feature	/note="match: GSS: Em:A0678815" 18389. .18823
misc_feature	/note="match: GSS: Em:A0270793" 25588. .26045
misc_feature	/note="match: GSS: Em:A0214655" 25694. .25967
misc_feature	/note="match: GSS: Em:A0097494" complement(41087. .41505)
misc_feature	/note="match: GSS: Em:A0213363" complement(41123. .41505)
misc_feature	/note="match: GSS: Em:A0133445" 45297. .45591
misc_feature	/note="match: GSS: Em:AF046498" 49180. .49534
misc_feature	/note="single clone region. short insert library only" complement(57981. .58764)
misc_feature	/note="match: GSS: Em:AF158011" complement(60582. .61043)
misc_feature	/note="match: GSS: Em:A0695221" 64797. .64841
misc_feature	/note="single clone region. short insert library only" 68250. .68620
misc_feature	/note="match: STS: Em:G16281" 70460. .70622
misc_feature	/note="match: STS: Em:G21080" complement(74554. .74895)
misc_feature	/note="match: GSS: Em:A0097224" complement(79470. .79747)
misc_feature	/note="match: STS: Em:AF191963" complement(79556. .79740)
misc_feature	/note="match: GSS: Em:A0541496" complement(79935. .80319)
misc_feature	/note="match: STS: Em:HSB052WGI" complement(85942. .86372)
misc_feature	/note="match: STS: Em:G22706" complement(86963. .87322)
misc_feature	/note="match: GSS: Em:A0191756" 87339. .87808
misc_feature	/note="match: GSS: Em:A0698012" 88780. .89285
misc_feature	/note="match: GSS: Em:A0121434" 94030. .94434
misc_feature	/note="match: GSS: Em:B89150" 94030. .94405

misc_feature /note="match: GSS: Em:AQ008338"
94054..94464
/note="match: GSS: Em:AQ010380"
join(112106..112211,116809..116933,123047..123132)
misc_feature /note="match: GSS: Em:AQ634365"
complement(join(112121..112215,116809..116936))
misc_feature /note="match: GSS: Em:AQ202845"
112126..112224
/note="match: GSS: Em:AF046385"
join(112145..112211,116809..116936)
misc_feature /note="match: GSS: Em:AQ634370"
complement(join(112150..112211,116809..116933,
123047..123172))
misc_feature /note="match: GSS: Em:AQ489599"
complement(113100..113712)
/note="match: GSS: Em:AQ477048"
complement(113175..113715)
/note="match: GSS: Em:AQ476615"
113710..114090
/note="match: GSS: Em:AQ434236"
join(116809..116933,123047..123119)
/note="match: GSS: Em:AQ732842"
116809..116941
/note="match: GSS: Em:AQ357462"
116809..116936
/note="match: GSS: Em:AQ047297 Em:AQ351276"
join(116840..116933,123047..123191)
/note="match: STR: Em:G13324 Em:G13474"
121507..121762
/note="match: GSS: Em:AQ482227"
complement(123025..123248)
misc_feature /note="match: GSS: Em:AQ779080"
123142..123248
/note="match: GSS: Em:AQ634365"
123192..123356
/note="match: STR: Em:G11177"
124453..124584
/note="match: GSS: Em:AQ002260"
complement(124491..124832)
misc_feature /note="match: STR: Em:G06939"
129943..130018
/note="single clone region"
144105..144499
/note="match: GSS: Em:AQ219599"
144105..144483
/note="match: GSS: Em:AQ231956"
complement(148672..149104)
misc_feature /note="match: GSS: Em:AQ672446"
152031..152510
/note="match: GSS: Em:AQ234353"
/note="match: GSS: Em:AQ234353"

ORIGIN
Query Match 3.9%; Score 60; DB 9; Length 166518;
Best Local Similarity 65.9%; Pred. No. 0.07;
Matches 87; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 533 GGACGAGTGTGCGGTGTGCGGAGCGGAGGCTCATCTGCTGTGACGGCTGCCCTCG 592
DB 160874 GGAGTCTCTCCGCGTGTGCAAGGACGGGGCGAGTCTGTGCGACGCGCTGCCCTC 160933
QY 593 GGGCTTCCACCTGCGCTGTGCTCCCTCCGCGGAGATCCCAAGTGGAGACCTGGAG 652
DB 160934 CTCCTACCACTGCACTTGTCTCAACCGCGCTGCCCGAGATCCCAACGGTGAATGGCT 160993
QY 653 GTGCTCCAGCTG 664
DB 160994 CTGCCCGCGCTG 161005

RESULT 69
LOCUS AR383186 1053 bp mRNA linear PAT 18-DEC-2003
DEFINITION Sequence 1 from patent US 6610823.

ACCESSION AR383186
VERSION AR383186.1 GI:40092750
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1053)
Targoff,I.N. and Ge,Q.
Antigens associated with polymyositis and with dermatomyositis
JOURNAL Patent: US 6610823-A 1 26-AUG-2003;
FEATURES
source 1..1053
Location/Qualifiers
1..1053
/mol_type="mRNA"

ORIGIN
Query Match 3.9%; Score 59.8; DB 6; Length 1053;
Best Local Similarity 58.9%; Pred. No. 0.17; 72; Indels 0; Gaps 0;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTGCGGTGTGCGGACGGGAGGCTCATCTGCTGTGACGGCTGCCCTC 591
DB 248 AGGATTACTGTGAGGTGTGCGGACGAGGAGTATTTCTGTGTGACACTGCCCTC 307
QY 592 GGGCTTCCACCTGCGCTGTGCTCCCTCCGCTCCGCGGAGATCCCACTGGACCTGGA 651
DB 308 GTGCTTCCACCTGCGCTGTGCTCCCTGATCTGACCTTGACCGGGCTCCAGAGGGCAATGGA 367
QY 652 GGTGCTCCAGCTGCTGCTGCGGACGAGTCCAGAGAGGTGACGCCCGCGGAGAGA 706
DB 368 GCTGCCCTCACTGTGAGAGAGAGGGGCTCACTGGAGGCCAAGAGAGAGAGA 422

RESULT 70
HSU08379
LOCUS HSU08379 1591 bp mRNA linear PRI 26-OCT-1995
DEFINITION Human M1-2 autoantigen 240 kDa protein mRNA, partial cds.
ACCESSION U08379
VERSION U08379.1 GI:761717
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1591)
Ge,Q., Nilasena,D.S., O'Brien,C.A., Frank,M.B. and Targoff,I.N.
Molecular analysis of a major antigenic region of the 240-kD
protein of M1-2 autoantigen
J. Clin. Invest. 96 (4), 1730-1737 (1995)
JOURNAL
MEDLINE 96013633
PUBMED 7560064
REFERENCE 2 (bases 1 to 1591)
Frank,M.B.
AUTHORS Direct Submission
TITLE Submitted (06-APR-1994) M. Bart Frank, Oklahoma Medical Research
Foundation, Arthritis and Immunology Program, 825 NE 13th Street,
Oklahoma City, OK 73104, USA
FEATURES
source 1..1591
Location/Qualifiers
1..1591
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="C4"
/cell_type="thymocyte"
/tissue_type="thymus"
/clone_lib="thymocyte and 5' stretch thymocyte libraries
(Clontech)"
1..>1591
/note="Northern blot with this cDNA shows hybridization of
a single band in 7.5 kb region"
/evidence="experimental"
1..>1591
/note="The partial cDNA encodes an antigenic region of a

nuclear protein to which autoantibodies are produced by 15-20% of dermatomyositis patients"

/codon_start=2
/product="M1-2 autoantigen 240 kDa protein"
/protein_id="AAC50228.1"
/db_xref="GI:761718"
/translation="PHFOOKQKQVEQKSSATLLTWGLBDEVHVFSEEDYHTLTNYKA
FQFMPLIAKKNPKIPIMSKMTTILGAKREFSANNPFKGSAAVAALAAAAAAYAE
OVSAAVSSATPIAPSGPALPPPADIDPPRIAPKTEGGPGHKRSKSPVDPG
RKLKRGKRAPIKILGILGKRGKSSDDEGPRAESDDLSGSVHASRPPGPV
RTKLKRGKRGKRVLCGPVAVGSEEVGYETIDQYCEVCQOQGEIILCDTCRA
YHLVCLDPELDRAPEGKSCPHCEKEGVMEAKEEBEYEEBEEBEEBDDHNEY
CEVCKDGGELICDCAICISYHICLNPPLDIPNGEWLCPKCTPVLKGRVKKILHMR
WCEPVAVAPQADGNPPVPPRPLOGSESEFEVYKWTGLSYMCSMAKEOLETFH
LVMRYNORKNMDPEPLDYGSGEDDGSDRKRYVDPHYAEMEEKYVAFGIKPEKMT
VRIITHNSDKKGNHYLVKGR"

ORIGIN

Query Match 3.9%; Score 59.8; DB 9; Length 1591;
Best Local Similarity 58.9%; Pred. No. 0.16;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGTCGGGACGGCGGGAGCTCATCTGCTGTGACGGCTGCCCTC 591
|||||
DB 765 AGGATTACTGTGAGTGTGTCAGCAGGCGTGGGAAATTATTCTGTGTGACACCTGCCCTC 824
|||||
QY 592 GGGCCTTCACCTGGCTGCTGCTCCCGCTCCGGGAGATCCCGAGTGGGACCTGGA 651
|||||
DB 825 GTGCTTACCACTCGCTGCTGTGATCTGAGCTTGAACGGGCTCCAGAGGGCAAAATGGA 884
|||||
QY 652 GGTGCTCCACCTGCTGCAAGCAACAGTCCAGAGAGGTGACGCCCGGCGAGAGA 706
|||||
DB 885 GCTGCCCTCACTGTGAGGAAGAGGGGATCCAGTGGAGGCAAGAGGAAGAAGA 939
|||||

RESULT 71
LOCUS CQ491372 6331 bp DNA linear PAT 30-JAN-2004
DEFINITION Sequence 23239 from Patent WO0160860.
ACCESSION CQ491372
VERSION CQ491372.1 GI:41456991
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Schlegel,R., Endege,W.O. and Monahan,J.E.
TITLE Genes differentially expressed in human prostate cancer and their
use
JOURNAL Patent: WO 0160860-A 23239 23-AUG-2001;
Milleium Predictive Medicine, Inc. (US)
FEATURES
source Location/Qualifiers
1..6331
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 3.9%; Score 59.8; DB 6; Length 6331;
Best Local Similarity 58.9%; Pred. No. 0.13;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGTCGGGACGGCGGGAGCTCATCTGCTGTGACGGCTGCCCTC 591
|||||
DB 1346 AGGATTACTGTGAGTGTGTCAGCAGGCGTGGGAAATTATTCTGTGTGACACCTGCCCTC 1405
|||||
QY 592 GGGCCTTCACCTGGCTGCTGCTCCCGCTCCGGGAGATCCCGAGTGGGACCTGGA 651
|||||
DB 1406 GTGCTTACCACTCGCTGCTGTGATCTGAGCTTGAACGGGCTCCAGAGGGCAAAATGGA 1465
|||||
QY 652 GGTGCTCCACCTGCTGCAAGCAACAGTCCAGAGAGGTGACGCCCGGCGAGAGA 706
|||||

DB 1466 GCTGCCCTCACTGTGAGGAAGAGGGGATCCAGTGGAGGCCAAGAGGAAGAAGA 1520
|||||

RESULT 72
LOCUS AX924075 6331 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 29 from Patent WO03080105.
ACCESSION AX924075
VERSION AX924075.1 GI:40217079
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Al-Mahmood,S., Colin,S. and Schneider,C.
TITLE Angiogenesis regulator genes, pharmaceutical preparations
containing same and uses thereof
JOURNAL Patent: WO 03080105-A 29 02-OCT-2003;
Gene Signal (FR)
FEATURES
source Location/Qualifiers
1..6331
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 3.9%; Score 59.8; DB 6; Length 6331;
Best Local Similarity 58.9%; Pred. No. 0.13;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGTCGGGACGGCGGGAGCTCATCTGCTGTGACGGCTGCCCTC 591
|||||
DB 1346 AGGATTACTGTGAGTGTGTCAGCAGGCGTGGGAAATTATTCTGTGTGACACCTGCCCTC 1405
|||||
QY 592 GGGCCTTCACCTGGCTGCTGCTCCCGCTCCGGGAGATCCCGAGTGGGACCTGGA 651
|||||
DB 1406 GTGCTTACCACTCGCTGCTGTGATCTGAGCTTGAACGGGCTCCAGAGGGCAAAATGGA 1465
|||||
QY 652 GGTGCTCCACCTGCTGCAAGCAACAGTCCAGAGAGGTGACGCCCGGCGAGAGA 706
|||||
DB 1466 GCTGCCCTCACTGTGAGGAAGAGGGGATCCAGTGGAGGCCAAGAGGAAGAAGA 1520
|||||

RESULT 73
LOCUS AX924076 6331 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 30 from Patent WO03080105.
ACCESSION AX924076
VERSION AX924076.1 GI:40217080
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Al-Mahmood,S., Colin,S. and Schneider,C.
TITLE Angiogenesis regulator genes, pharmaceutical preparations
containing same and uses thereof
JOURNAL Patent: WO 03080105-A 30 02-OCT-2003;
Gene Signal (FR)
FEATURES
source Location/Qualifiers
1..6331
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 3.9%; Score 59.8; DB 6; Length 6331;
Best Local Similarity 58.9%; Pred. No. 0.13;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGTCGGGACGGCGGGAGCTCATCTGCTGTGACGGCTGCCCTC 591
|||||

```

Db      1346 AGGATTACTGTGAGGTTGGCCACAGGTTGGGAAATTTATCTGTGTGACACTGCGCTTC 1405
Qy      592 GGGCCTTCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGAGGACCTGGA 651
Db      1406 GTGGCTACACACCTCGTCTGCTTGTGATCTGAGCTTGACCGGGCTCCAGAGGGCAATGGA 1465
Qy      652 GTGCTCCAGCTGCTGCTGACGCAACAGTCCAGAGGTGACGCCCGGGCAGAGGA 706
Db      1466 GTGGCTCCAGCTGCTGAGGAGGAGGGGCTCCAGTGGAGGCCAAGAGGAAGA 1520

RESULT 74
AF006515 6331 bp mRNA linear PRI 27-NOV-1997
LOCUS Homo sapiens CHD3 mRNA, complete cds.
ACCESSION AF006515
VERSION AF006515.1 GI:2645432
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 6331)
Moodage, T., Basrai, M.A., Bayevanis, A.D., Hietzer, P. and Collins, F.S.
Characterization of the CHD family of proteins
Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)
9326634
PUBMED 9326634
2 (bases 1 to 6331)
Moodage, T.
Direct Submision
Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human
Genome Research Institute, National Institutes of Health, 49
Convent Drive, Bethesda, MD 20892-4442, USA
Location/Qualifiers
1..6331
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p13; between WI-9178 and D17S786"
1..6331
/gene="CHD3"
211..6045
/gene="CHD3"
/codon_start=1
/product="CHD3"
/protein_id="AAB87383.1"
/db_xref="GI:2645432"
/translation="MKAADTVILMARSKNDQLRISFPGLCWGRMPDKDRIILLPSA
IGVKKRGRPKQKQENKPKRKRKDEEFGSGRDEYREKSGSEYGTGPBRK
RRKIKREKKTGRKKGEGDGGQKOVOKSASATLLTWGLDEYHVSSEYHTLT
NYKAFQFMRPLIAKNPKTIPMSKMTTIGAKWRSEANPFGSAAVAATAAAAA
AAVQVSAVSSATDPIAASGPALPPPPADLOPPIRAKTKEGSGPHKRSKSP
VVDGRKRLGKMAPIKTLGLIGKRRKKGSTVPSGDEPERAESDLDGSHSA
SRPQGVPTKTLKRGPRGRKKKVLGCAVAGEEVDYETDHDQYCEVCGQGGII
LEDTCPRATHVCLDPBELDRABEGKWSCHCEKEGQWAKSEEEYEEGEBEKE
BEDDMEYRCVCDGELLCDACISSYHICNPLPDIPNGEMLCPSCTPLVSAFLTI
VOKILHWGGEPPVAPAPQADGNPDVPPRPLQGRSREPPKVGWGLSYMSAK
ELQLEIFHLVWYRNORKNDBRPPLDYGSDDDGSKRKYKVDHVAEMEKYFR
GIKPRMTYHRIINNSVDKKNYHLYVKRDLPYDOSTWEEBMANIPRYEHNKOSTYR
HRELNGEDPPAPRKYKKKKELQGDGPPSPPTNDYVYETDPRITVATGGLHWY
LEGILWLRFSMAQGTDTLLADEGLQKTQITVFLSLYKGTGKPLVSAFLTI
NMRERQWMAKPYVVTYTGDKDSRAIIRENEFSFDNAIKGGKAFKKKRAOVFV
VLTSEYELITDOALGSIKMACLVDEDAHRLKNQSKFRLVNGYKIDHKLLTGP
LONNLEELHLNPLFPERPNNLGFLSEFADISKEDQIKLHDLIGPMILRLKADV
FKMNAKTELIVRVELSPMKCYKYIILRNPEALNSRGAGVSLINIMMLKCN
HYLPVPAAMESPKLPSGAYEGGALIKSGKMLLQKMLRKLKEGSHRVLITSQNTX
LULLEDFLDYEGKYERIDGGITGALRQALIDRFNAPGAQCFLLSTAGGGINLA
7ADVTIIFSDMNPNDIOAFRAHRIQGANVMYRPTVASVERITOVAKRKML
THLVVPGIGSKAGSKQELDLDFGLTEELFKDNEGENKEDSGVIHYNEALAR
LIDRNDATEDPTDVQNMNEYLSFKVAQYVVEBEDTIEIERIILIQEENVDPDWEK

```

```

LIRHYEQOQEDLARNLKGKRVKQVNTNDAAQEDQDNQSEYVSGSEBEDPDERP
EGRSKQRLNREKPLPLARVGNIEVGFNTROKATLNAWRKMPQDAFT
TQWLVRDLNGKTEKEFKAVSLFMRHLCEPADGSETFADGVRELSQVLTIRGV
MSLVKKQVQEFTEHNGRWSMPLMELPADSRRASPTPTSPTTSPASATNSCT
SKPATPASEKEGIRTPLEKEANQOEKPKNSRIIGKMETEADAPSPASIGRL
EBPKTPLDEBVGVPGBMEBERGVGDGKSDVDGDLRLRGPDPEBPSNRBEKT
EKPRFMNADGFTIELHLMONERRAIISSKKNIEIWRBDYLLAGIVLHGVARW
QDIQNDQFATINBPKTEBANKFLNEMKNFLARFKLEQALYTEODRLDAVYLN
SQEPAPPAVALARPAEACLAESHQLSKESLAKNKAPNAVLAHLVNLQLEBLSMK
ADVTLPATLSRIPIIARLQWSEKSIISRLASKTEBPHPTPAPYPPGYPATPGYAA
PSAIPVGAALAAAGVNSQWPAGSFITAATNGPIHRFKKVPEIQ"

ORIGIN
Query Match 3.9%; Score 59.8; DB 9; Length 6331;
Best Local Similarity 58.9%; Pred. No. 0.13;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy      532 AGGACAGTGTGACCGTGTGTCGGGACCGGGGAGTCTATCTGTGTGACGCTGCGCTTC 591
Db      1346 AGGATTACTGTGAGGTTGGCCACAGGTTGGGAAATTTATCTGTGTGACACTGCGCTTC 1405
Qy      592 GGGCCTTCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGAGGACCTGGA 651
Db      1406 GTGGCTACACACCTCGTCTGCTTGTGATCTGAGCTTGACCGGGCTCCAGAGGGCAATGGA 1465
Qy      652 GTGCTCCAGCTGCTGCTGACGCAACAGTCCAGAGGTGACGCCCGGGCAGAGGA 706
Db      1466 GTGGCTCCAGCTGCTGAGGAGGAGGGGCTCCAGTGGAGGCCAAGAGGAAGA 1520

RESULT 75
AX924049 6771 bp DNA linear PAT 18-DEC-2003
LOCUS Sequence 3 from Patent WO03080105.
DEFINITION AX924049
ACCESSION AX924049
VERSION AX924049.1 GI:40217057
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Al-Mahmoud, S., Collin, S. and Schneider, C.
Angiogenesis: regulator genes, pharmaceutical preparations
containing same and uses thereof
Patent: WO 03080105-A 3 02-OCT-2003;
JOURNAL Gene Signal (PR)
FEATURES
Location/Qualifiers
1..6771
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 3.9%; Score 59.8; DB 6; Length 6771;
Best Local Similarity 58.9%; Pred. No. 0.13;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy      532 AGGACAGTGTGACCGTGTGTCGGGACCGGGGAGTCTATCTGTGTGACGCTGCGCTTC 591
Db      1286 AGGATTACTGTGAGGTTGGCCACAGGTTGGGAAATTTATCTGTGTGACACTGCGCTTC 1345
Qy      592 GGGCCTTCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGAGGACCTGGA 651
Db      1346 GTGGCTACACACCTCGTCTGCTTGTGATCTGAGCTTGACCGGGCTCCAGAGGGCAATGGA 1405
Qy      652 GTGCTCCAGCTGCTGCTGACGCAACAGTCCAGAGGTGACGCCCGGGCAGAGGA 706
Db      1406 GTGGCTACACACCTCGTCTGCTTGTGATCTGAGCTTGACCGGGCTCCAGAGGGCAATGGA 1465

RESULT 76
HSU91543

```

LOCUS HSU91543 6771 bp mRNA linear PRI 04-AUG-1998
DEFINITION Homo sapiens zinc-finger helicase (hzfh) mRNA, complete cds.
ACCESSION U91543
VERSION U91543.1 GI:3298561
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Aubry, F., Matei, M.G. and Galibert, F.
TITLE Identification of a human 17p-located cDNA encoding a protein of
the Snf2-like helicase family
JOURNAL Eur. J. Biochem. 254 (3), 558-564 (1998)
MEDLINE 98351552
PubMed 9888266
REFERENCE 2 (bases 1 to 6771)
AUTHORS Aubry, F.
TITLE Direct Submision
JOURNAL Submitted (13-DEC-1996) UPR41, CNRS, 2, Avenue Leon Bernard, RENNES
35000, France
FEATURES
source Location/Qualifiers
1..6771
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p12-13"
/cell_line="Hela"
<1..>6771
/gene="hzfh"
151..6153
/gene="hzfh"
/note="Snf2-like helicase family"
/codon_start=1
/product="zinc-finger helicase"
/protein_id="AAC39923.1"
/db_xref="GI:3298562"
/translation="MKAADTVILMAASKNDOLRISPPGICNDPMDDKIDRLIPSA
LGVKKRGRPKQKQENKPKPKRKRKRDSEBERGSRDEREKSSEGGSEYGTGPKRK
RRKRERKEKTKRRKKKGEGQKQVQKSSATLLTWLEDVHVSSEDTYHLLT
NYKAFQFMRPLAKNPKIPMSKMMTILGAKRREPSANNPFSGAAVAATAA
AAVAGQSAVSSATPIAPSGPPALPPPPADLOPPPIRAPKTEGKPGHKRSKSPR
VDFGRKLGKKMAPIKILGLGKRRKGSYFQSDGDBREABESDLDSGVHSA
SRPGSPVATKLRGPRGRKKKYLGCSCVCHCEKSGQVMEAKEBEEYEEGEGEKE
LEDTCRAHYILCDPELDRAFBGKSGCHCEKSGQVMEAKEBEEYEEGEGEKE
EEDDHMEYCRVCDGSELCCDACSISYIHCLNPLPIPIPMEMLCPRCTCPVLKGR
VOKILMRWGEPPVAAPQAQADNPDPVPPRLQGRSREFFKVVGVCSYMSKAK
ELQLEIFHLVMRYNQKNDMDPEPLDYGSGEDGSKDKRKYDHPVAMEKRYRF
GKPEMWTYHRIINHSVDKKGNHYLVKMRDLPYDQSTWSEEDMNIPEYEEKOSYMR
HRELIMGEDPADOPRYKKKKKELCGDGPSSPTNPDTVYETQPRITANGGTLHYO
LGGANLRSMAQGTDTTILADENGAKTITQTVLYSLYKEGHTKPPVLSAPLSITL
MNEREQMAPRYVATVYTGDKDSRAIIRENEFSFEDNAIKGKGAFFKKRREQVFF
VLTYSYELITTDQALGSIKMACLVVDEAHARKNNQSKFRVLYNGKIDHKLITGP
LQNNLEELFHLINFLTPERFMNKEGLFEFAFADISKDQIKLHDILGPMRLRKADV
FKMNPKEELIVLSELPFPMQKKYKYLIRNFPALNSRGGOVSLNTIMMDKKCN
HYLPVPAAMESPKLPSGAVGGALIKSSGKMLQKMLRKLEQGHRLVLSOMTKM
LHLBEPFLDYEGYKRYRIDGIGITGALROAIDRPAKPAQOQCFILSTAGGIGITLA
TADTVLIFDSMDMHNIDIOAFSAHRIQANKWMTIRFVTRASVERKITQVARKMML
THLVAPGSGSKAGSSKQELDDILKFTGEEFKQENBENKEEDSSVLIHYNEALR
LIDRNDATEDETDVQNMNELYLSFKAQVYVEEDKIEIEREIIQEBNVDPYWEK
LIRHYEQOQEDLARILGKGRVROVNTNDAQEDQNDQSESVSGSEDEDFERP
EGRROSKRQDLRNEKDELPIILARVGNTEVIGFNTROKKAFLINAMRWGMPODAFT
TQWLVADLAKGTEKEKAYVSLFMRILCERGADGSETPDGVPRELSQOYITRIGV
MSLVKKVQEFETINGRWMPBELMPPSADSKRSAPSPKTSPTTPASATNSCT
SKPATPAQVEGEGKITPLEKEANQEKPEKNSRIIGKMETEADAPAPASIGRL
BRKTIPLDEVPVPGEMPEPEYGRDRKSAKTESYTGREGKPEKQDGEHMERPE
TQDLGREGDYKGRLELPGPRDPRSGREKTEENPRFMPNLADGDFELHTLQNE
EBAATSSGKLEIWHRRHLYMILAGIVLHGVARMODIONDAQPAALHARAEKCLAS
ELEMKKRPLARPEKILEQALVIEQJARAAYINTLSEPRAPMAALHARAEKCLAS
HQHLSKESLAKNKANAVLHKVLNQLEBLISDKKAVTLPATLSLITPTLARLOQSB
RILSLRLASGTEPHPTPAVPPGYPATPPGYGAFAFAAPVGAALAAAGANYSQMPAGSF
IIPAATGPPVLVKEKEKEMVGAIVSDGLDRKEPRAGEVICIDP"

misc_feature join(151..5074,5177..6153)
/gene="hzfh"
/note="encodes putative zinc-finger helicase formed from
alternative splicing; mRNA observed, but no protein yet
identified; hzfh2"
misc_feature 1294..1419
/gene="hzfh"
/note="encodes zinc-finger region"
misc_feature 1525..1650
/gene="hzfh"
/note="encodes zinc-finger region"
misc_feature 2239..2259
/note="encodes zinc-finger region"
misc_feature 2419..3678
/gene="hzfh"
/note="encodes ATPase/helicase domain"
misc_feature 3010..3060
/gene="hzfh"
/note="encodes nuclear targeting sequence"
polyA_site 6744..6749
/gene="hzfh"
/note="polyA_site"

ORIGIN

Query Match 3.9%; Score 59.8; DB 9; Length 6771;
Best local Similarity 58.9%; Pred. No. 0.13; Mismatches 0; Gaps 0;

Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 513 AGGACGAGTGTCCGTGTGCGGACGCGGAGACTATCTGTGTGACGCGTGCCTC 591

Db 1286 AGGATTACTGTGAGGTGTCCAGAGGGTGGGAATATTCTGTGTGACACTGCGCTC 1345

Qy 592 GGGCTTCCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 651

Db 1346 GTGCTTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1405

Qy 652 GGGCTTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 706

Db 1406 GCTGCTTCTACTGTGTGAGAGAGGGGCTTCAATGTGGAGGCCAAGAGAGAGAGA 1460

RESULT 77

AC135674/c 168479 bp DNA linear HTG 15-NOV-2002

LOCUS Rattus norvegicus clone CH230-368023, WORKING DRAFT SEQUENCE.

DEFINITION AC135674

ACCESSION AC135674.2 GI:25007200

VERSION HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 168479)

Muzny, D. Marie, Metzker, M. Lee, J. Adamson, S., Adams, C., Alder, J.,

Allan, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Behnmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Buttrill, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, R., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dedertich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, N., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Becotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregregios, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Huylk, S., Hume, J., Idelbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karchay, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshew, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martineau, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Nandasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelamen, O., Okunolu, G., Olarnunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, M., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.U., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valse, R., Vere, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Wortley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
Direct Submision
Unpublished

JOURNAL
2 (bases 1 to 168479)

REFERENCE
Rat Genome Sequencing Consortium.

AUTHORS
Submitted (21-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
3 (bases 1 to 168479)

REFERENCE
Rat Genome Sequencing Consortium.

AUTHORS
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
On Nov 15, 2002 this sequence version replaced gi:24181980. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: KCMV
Center clone name: CH230-368023

----- Summary Statistics
Assembly program: Phrap, version 0.990329
Consensus quality: 163230 bases at least Q40
Consensus quality: 165491 bases at least Q30
Consensus quality: 166611 bases at least Q20
Estimated insert size: 169544; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
1 168479: contig of 168479 bp in length.

FEATURES
source
Location/Qualifiers
1..168479
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-368023"
complement(2776..3643)
/note="clone_boundary"
clone_end:77
site:
end_sequence:B2145823"
complement(166515..167327)
/note="clone_boundary"
clone_end:5p6
site:
end_sequence:B2145824"

ORIGIN
Query Match 3.9%; Score 59.8; DB 2; Length 168479;
Best Local Similarity 55.6%; Pred. No. 0.076;
Matches 115; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 533 GGACGAGTGTGCGGTGTGTGCGGAGCGGAGGAGCTCATCTGTGAGCGGCTCGCG 592
DB 12557 GAGATTTCTCCGCGTGTGCAAGACGATGCGAGCTTCTGTGTGATGCTTTCTCC 12498
QY 593 GGCCTTCCACCTGCGCTGTGCTCTCCCTCCGCTCGGAGATGCCAGTGGACCTGAG 652
DB 12497 CTCCTACCACTTGCACTGCTCTCAACCGCGCTGCGGAGATCCGACGATGATGCT 12438
QY 653 GTCTCTACCTGCTCTGAGCAACATCTCCAGAGGTGCGAGCGGAGAGAGCGCCG 712
DB 12437 CTGCGCGCGCTCTCAAGTATGTTACACTCTCTGCGCGCGCGCGCGCGCG 12378
QY 713 GCCCGAGAGCCAGCCGCTGAGAGCCCG 739
DB 12377 CCG 12351

RESULT 78
AC119782/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-134W24, WORKING DRAFT SEQUENCE.
ACCESSION AC119782
VERSION AC119782.5 GI:23670853
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
1 (bases 1 to 221787)
Muzny, D., Marie, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Behnmed, F., Biwalta, K., Blait, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

RESULT 84
 LOCUS D86113 63 bp DNA linear PRI 04-JUL-1997
 DEFINITION Homo sapiens gene for HC21EXC33, exon.
 ACCESSION D86113
 VERSION D86113.1 GI:2244722
 KEYWORDS HC21EXC33.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Kudoh,J., Nagamine,K., Asakawa,S., Abe,I., Kawaaki,K., Maeda,H., Tsujimoto,S., Minoshima,S., Ito,F. and Shimizu,N.
 TITLE Localization of 16 exons to a 450-kb region involved in the autoimmune polyglandular disease type I (APECED) on human chromosome 21q22.3
 JOURNAL DNA Res. 4 (1), 45-52 (1997)
 MEDLINE 97333005
 PUBMED 9179495
 REFERENCE 2 (sites)
 AUTHORS Nagamine,K., Kudoh,J., Asakawa,S., Abe,I., Maeda,H., Tsujimoto,S., Minoshima,S., Ito,F. and Shimizu,N.
 TITLE Localization of 22 Exons to a 450-kb Region Involved in the Autoimmune Polyglandular Disease Type I (APECED) on chromosome 21q22.3
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 63)
 AUTHORS Shimizu,N.
 TITLE Direct Submission
 JOURNAL Submitted (19-JUN-1996) Nobuyoshi Shimizu, Keio University School of Medicine, Department of Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@mb.med.keio.ac.jp, Tel:03-3351-2370, Fax:03-3351-2370)
 FEATURES
 source
 1..63
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.3"
 /cell_line="GM130B"
 /cell_type="B-lymphoblastoid"
 /clone_11b="chromosome 21-specific cosmid library"
 1..63
 /gene="HC21EXC33"
 1..63
 /gene="HC21EXC33"
 ORIGIN
 Query Match 3.8%; Score 58.2; DB 9; Length 63;
 Best Local Similarity 95.2%; Pred. No. 0.56;
 Matches 60; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1149 GATGACATGCGCAGCAGCGCCGCTGTGCACAGGAGTACCTGGAGTCCCTTGAGC 1208
 |||
 1 GATTGACACCGCAGTCAAGCGCCGCTGTGCACAGGAGTGTGAGTCCCTTGAGC 60
 Db 1209 GAG 1211
 |||
 61 GAG 63
 RESULT 85
 LOCUS BC060721 4343 bp mRNA linear ROD 25-MAY-2004
 DEFINITION Mus musculus chromodomain helicase DNA binding protein 3, mRNA
 ACCESSION BC060721
 VERSION BC060721.1 GI:38511411
 KEYWORDS

SOURCE
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 4343)
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,K.H., Scheffer,C.F., Bhat,N.K., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stabileton,M., Soares,M.B., Bonaldo,M.F., Cavaent,T.L., Scheetz,T.E., Brownstein,M.J., Uesdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S., Loquailano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Vialalon,D.K., Wuzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Hellon,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schultz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalka,U., Smalins,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 PUBMED 12477932
 REFERENCE 2 (bases 1 to 4343)
 AUTHORS Strausberg,R.
 TITLE Direct Submission
 JOURNAL Submitted (03-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-rc@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: M. Benito Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mc@paxll.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 125 Row: f Column: 19
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.
 FEATURES
 source
 1..4343
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6406263"
 /tissue_type="Brain, enriched mouse brain 12.5dp"
 /clone_11b="NIH BMAP_F00"
 /lab_host="DH10B"
 /note="Vector: pYX-A5C"
 ORIGIN
 Query Match 3.8%; Score 58.2; DB 10; Length 4343;
 Best Local Similarity 58.3%; Pred. No. 0.28;
 Matches 102; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
 Oy 532 AGGACGAGTGTGCGGTGTGTGGGAGCGCGGAGAGCTATCTGCTGACGCGCTCCCTC 591
 |||
 512 AGGATTACTGTGAGGTGTGCGCAGCGCGGCGGGAATATTCTGTGCGCAGCTGCCCC 571

QY 592 GGGCCCTCCAGCTGCGCTGCTCCCTCCGTCGGAGATCCCAAGTGGAGCACTGGA 651
AC133595/LOCUS 116984 bp DNA 1linear HTG 27-MAR-2003
DEFINITION Homo sapiens chromosome 17 clone RP11-242A8 map 17, 3 unordered
pieces.
AC133595 GI:27452934
VERSION HTG: HTGS PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 116984)
AUTHORS Birren, B., Nussbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 17, clone RP11-242A8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 116984)
AUTHORS Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Basleten, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camarrato, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand, Piere, N., Hago, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Jones, C., Kamat, A.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneu, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Piere, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tessie, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
TITLE Submitted (14-SEP-2002) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 116984)
AUTHORS Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Basleten, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camarrato, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand, Piere, N., Hago, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Jones, C., Kamat, A.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneu, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Piere, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tessie, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
TITLE Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jan 1, 2003 this sequence version replaced gi:27369450.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L28162
Center clone name: 242_A_8

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 40901: contig of 40901 bp in length
* 40902 41001: gap of 100 bp
* 41002 78861: contig of 37860 bp in length
* 78862 78961: gap of 100 bp
* 78962 116984: contig of 38023 bp in length.
FEATURES
source
1..116984
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-242A8"
/clone_11b="RP11-11 Human Male BAC"
ORIGIN
Query Match 3.8%; Score 58.2; DB 2; Length 116984;
Best Local Similarity 59.3%; Pred. No. 0.17; Mismatches 99; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
Matches 99; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 532 AGGACAGTGTGCGCTGCTCGGAGACGGGAGACTATCTGCTGACGCGCTGC 591
AC104581 187718 bp DNA 1linear PRI 13-MAY-2003
LOCUS Homo sapiens chromosome 17, clone RP11-1099W24, complete sequence.
DEFINITION AC104581 GI:30581670
VERSION HTG:
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 187718)
AUTHORS Birren, B., Nussbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 17, clone RP11-1099W24
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187718)
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Basleten, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarrato, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, A., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lenoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPeeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, W., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trillio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (13-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 187718)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (10-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 187718)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Baetien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22268
Center clone name: 1099_M_24

FEATURES	source
Location/Qualifiers	1..187718
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="17"	
/map="17"	
/clone="RP11-1099M24"	
/clone_11b="RP11 Human Male BAC"	
789..1020	
/rpt_family="MIR"	
1539..1582	
/rpt_family="CGA"	
1675..1676	
/note="unresolved homopolymeric run: restriction enzyme digest fingerprint indicates about bp of poly C repeat missing here."	
4209..4468	
/rpt_family="MIR"	
4732..4902	
/rpt_family="CAGAGA"	
complement(7246..7250)	
/note="30 qual SNGL region"	
10077..10135	
/rpt_family="CCA"	
10625..10654	
/rpt_family="CAG"	
10880..10977	
/rpt_family="C-rich"	
11178..11217	
/rpt_family="CAG"	
11775..11817	
/rpt_family="CCA"	
11836..11918	
/rpt_family="C-rich"	
16622..16642	
/rpt_family="GC-rich"	
16905..16934	
/rpt_family="AT-rich"	
complement(16991..17008)	
/note="30 qual SNGL region"	
17644..17674	
/rpt_family="AT-rich"	
18005..18020	
/rpt_family="AT-rich"	
complement(23383..23647)	
/rpt_family="AluUo"	
complement(23661..23811)	
/rpt_family="MIR"	
24851..25513	
/rpt_family="12"	
25633..26008	
/rpt_family="AluSg"	
complement(27601..27833)	
/rpt_family="AluUo"	
complement(27834..27858)	
/rpt_family="R1ggr4(Zombi)"	
27859..28148	
/rpt_family="AluSx"	
complement(28149..28277)	
/rpt_family="R1ggr4(Zombi)"	
complement(28432..28554)	
/rpt_family="MIR"	
complement(28635..28988)	
/rpt_family="MUTIG1"	
complement(29134..29251)	
repeat_region	

Center: Whitehead Institute/ MIT Center for Genome Research


```

repeat_region      /rpt_family="L2"
                    complement(29321, .29644)
                    /rpt_family="AluY"
repeat_region      29776, .30152
                    /rpt_family="L2"
                    complement(30196, .30508)
repeat_region      /rpt_family="L2"
                    30333, .30339
                    /note="1340 bp of E.coli insertion sequence in BAC excised
                    here"
repeat_region      complement(30509, .30560)
                    /rpt_family="L1MB7"
repeat_region      complement(30571, .30868)
                    /rpt_family="AluSg"
repeat_region      complement(30869, .31610)
                    /rpt_family="L2"
repeat_region      31831, .31926
                    /rpt_family="MIR"
repeat_region      complement(32394, .32578)
                    /rpt_family="MIR"
repeat_region      32717, .33023
                    /rpt_family="AluSx"
repeat_region      33350, .33522
                    /rpt_family="AluSg/X"
repeat_region      complement(33823, .33836)
                    /rpt_family="AluYo"
repeat_region      33837, .34142
                    /rpt_family="AluSx"
repeat_region      complement(34143, .34263)
                    /rpt_family="AluYo"
repeat_region      34455, .34640
                    /rpt_family="MIR"
                    complement(34899, .35191)
                    /rpt_family="AluSx"
repeat_region      /rpt_family="AluSx"

```

Query Match 3.8%; Score 58.2; DB 9; Length 187718;

Best Local Similarity 59.3%; Pred. NO. 0.15; 68; Indels 0; Gaps 0;

```

Matches 99; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 532 AGGACAGTGTGCTGCTGTGCGGAGCGCGGAGCTCATCTGTGACGCGCTGCCCTC 591
    |||||
DB 57712 AGGATTAAGTGTGAGTGTGCGACAGGCGGGAATAATTTCTGTGACACTGCCCTC 57771
    |||||

QY 592 GGGCCCTTCAACCTGCGCTGCTGCCCTGCCCTCCGCGGAGATCCCGAGTGGGACCTGGA 651
    |||||
DB 57772 GGGCCCTTCAACCTGCGCTGCTGCCCTGCCCTCCGCGGAGATCCCGAGTGGGACCTGGA 57831
    |||||

QY 652 GGTGCTCCAGCTGCGCTGCGGAGCAACAGTCCAGAGGAGGAGCGCCCGG 698
    |||||
DB 57832 GGTGCTCCAGCTGCGCTGCGGAGCAACAGTCCAGAGGAGGAGCGCCCGG 57878
    |||||

```

```

RESULT 88
AC11597      232984 bp   DNA      linear   HTG 09-NOV-2002
LOCUS      Rattus norvegicus clone CH230-150P11, WORKING DRAFT SEQUENCE, 3
DEFINITION      unnumbered pieces.
ACCESSION      AC11597
VERSION      AC11597.4 GI:24818936
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                Ratus.
REFERENCE      1 (bases 1 to 232984)
AUTHORS      Muzny,D,Marle,M, Lee,M, Abramson,S, Adams,C, Alder,J,
                Allen,C, Allen,H, Albrechts,S, Amin,A, Anguiano,D,
                Anyadebech,V, Aoyagi,A, Ayodeji,M, Baca,B, Baden,H,
                Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benhmed,F,
                Biswal,J, Blair,J, Blankenburg,K, Blych,P, Brown,M,
                Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
                Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A,

```

```

TITLE      JOURNAL
REFERENCE      JOURNAL
AUTHORS      JOURNAL
TITLE      JOURNAL
JOURNAL      JOURNAL
COMMENT      Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
                of Molecular and Human Genetics, Baylor College of Medicine, One
                Baylor Plaza, Houston, TX 77030, USA
                On Nov 9, 2002 this sequence version replaced gi:23264551.
                The sequence in this assembly is a combination of BAC based reads
                and whole genome shotgun sequencing reads assembled using Atlas
                (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
                in the feature table below represents a scaffold in the Atlas
                assembly (a 'contig-scaffold'). Within each contig-scaffold,
                individual sequence contigs are ordered and oriented, and separated
                by sized gaps filled with Ns to the estimated size. The sequence
                may extend beyond the ends of the clone and there may be sequence
                contigs within a contig-scaffold that consist entirely of whole
                genome shotgun sequence reads. Both end sequences and whole genome
                shotgun sequence only contigs will be indicated in the feature
                table.

```

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>

```

Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GMDP
Center clone name: CH230-150P11
-----
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 198152 bases at least Q40
Consensus quality: 199686 bases at least Q30
Consensus quality: 200691 bases at least Q20
Estimated insert size: 204955; sum-of-coverage: 7x in Q20 bases; sum-of-coverage: 7x in Q20 bases; sum-of-coverage: 7x in Q20 bases
-----
NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 26928: contig of 26928 bp in length
* 26929 27028: gap of unknown length
* 27029 231727: contig of 204699 bp in length
* 231728 231827: gap of unknown length
* 231828 232984: contig of 1157 bp in length.
Location/Qualifiers
1. 232984
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-150P11"
1. 1443
/notes="wgs_end_extension
clone_end:T7"
456. 843
/notes="clone_boundary
clone_end:T7
site:BCORI
end_sequence:BH33066"
ORIGIN
Query Match 3.7%; Score 57.6; DB 2; Length 232984;
Best Local Similarity 64.0%; Pred. No. 0.19; Indels 0; Gaps 0;
Matches 87; Conservative 0; Mismatches 49;
QY 529 ATGAGAGCAGTGTGCGGTGTGCGGACGCGGAGGCTCATCTGTGACGCGTGGC 588
DB 72989 ATATGGAATTCTGTGCAATGTGCAAGATGCGGAGAACTACTATGCTGTGACACTGCGC 73048
QY 589 CTCGGGCGCTTCCACCTGCGGCTGCTCCCTCCGCTCCGCGAGAGATCCCGACGTGGACCT 648
DB 73049 CTTCTTCTTACCAATTCACCTGTGTGAACCCCGCTGCGACAGATCCCAATGGCGAAT 73108
QY 649 GGAGGTGTCTCAGCTG 664
DB 73109 GGCTCTGTCCCGCTG 73124
-----
RESULT 89
AC115415 261384 bp DNA 1linear HTG 21-SEP-2002
AC115415 Rattus norvegicus clone CH230-164A18, *** SEQUENCING IN PROGRESS
DEFINITION *** 2 unordered pieces.
ACCESSION AC115415
VERSION AC115415.3 GI:23266199
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE
AUTHORS
1 (bases 1 to 261384)
Muzny, D. Marie, Metzker, M. Lee, Abrazon, S., Adame, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amun, A., Anguiano, D.,
Anyalebech, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Barnham, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Caldens, V., Carter, K., Cavazos, I., Cesar, H., Chen, A.,
Chen, J., Chavaz, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, L., Cree, A., D'Souza, L.,
Dalla, M. L., Davis, C., Davy-Carroll, L., De And, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Drapper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Hulik, S., Hume, J., Idledit, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensbaw, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Miklosavljewic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundaas, M., Murphy, M., Natr, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackelumb, O., Okumu, G., Olarunagbon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puazo, M., Qutroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajic, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, J., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Vales, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 261384)
Worley, K. C.
Direct Submission
Submitted (19-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 261384)
Baylor Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21737225.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence only
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
-----
Genome Center

```


DVQGAHAQPTLTQPSIQVQFHHQKTLKSHFOQOPOOQLPPPPPLPPPPPLPPPP
 OQHPPLPSPSOHLASQSHSPRPAPASOMDMHHEFELEMOLELLOAOBPLO
 LQOTSPHOLQOTIVQIVROPAPOVQSGOETLOATDEBPASQSGKRALPLDKN
 TTAALPQASGEPTLSPVPVDSTIQSSNNVVRKSTISIMFSITLMEISSTRLE
 RHEPQIQSVSNITRGAPOVPSLSPAPRNVSLTSSVONQAMPSTLTLQVPSLV
 HSTPQMPNLISDSQAMASLADHPQAPPSLMSGTQVPSLATPLOSIPPSVMO
 PRTGSSSSGRTSGSLCPRDGADPDLLENALCKRSPBEMLSKLPFLIGOKIKHWPYS
 DSLTVLRPLMTTMSSEFDSCHRFTEPRWKVRLPLASLTMYDKVP"

ORIGIN

Query Match 3.7%; Score 56.6; DB 9; Length 6505;
 Best Local Similarity 55.3%; Pred. No. 0.54;
 Matches 110; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 526 AGAATGAGGACGATGCGCGTGTGCGGAGCGCGGAGGACCTGCTGTGAGCGGCT 585
 DB 3352 AGAATGAGGACCTTCTGTGCTTCTCAATGCGGAGAGTTACTGTGTGACCGCT 3411
 QY 586 GCCCTCGGCGCTTCACTGCGCTGCTGCTCCGCTCCGCGAGATCCCGAGTGGGA 645
 DB 3412 GCCCAAGAGTTCACCTCTCTGCGATGTCGACGCTTGTGCTCCAGGGGGAG 3471
 QY 646 CCTGAGAGTCTCACTGCTGCGAGCAACACTCAGAGAGTGCAGCCCGGCGAGAG 705
 DB 3472 AGTGGGTGTGTACTTGTGCGGAGCCTGACCGAGCCGAGATGAGTACGACTGTGAGA 3531
 QY 706 AGCCCGCGCGCCGAGAGCC 724
 DB 3532 ATGCTGCTATACGAGCC 3550

RESULT 94
 LOCUS AF210315 5811 bp mRNA linear INV 31-MAY-2001
 DEFINITION Drosophila melanogaster bonus (bon) mRNA, complete cds.
 ACCESSION AF210315
 VERSION AF210315.1 GI:6630999
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 5811)
 Beckstead, R., Ortiz, J.A., Sanchez, C., Prokopenko, S.N., Chambon, P.,
 Losson, R. and Bellen, H.J.
 Bonus, a Drosophila homolog of TIF1 proteins, interacts with
 nuclear receptors and can inhibit betaFTZ-F1-dependent
 transcription
 JOURNAL Mol. Cell 7 (4), 753-765 (2001)
 MEDLINE 11336699
 PUBMED 21235528
 REFERENCE 2 (bases 1 to 5811)
 AUTHORS Beckstead, R.B., Prokopenko, S.N. and Bellen, H.J.
 TITLE Direct Submision
 JOURNAL Submitted (01-DEC-1999) Molecular and Human Genetics/HMTI, Baylor
 College of Medicine, T634, Mail Stop BCM235, One Baylor Plaza,
 Houston, TX 77030, USA
 FEATURES
 source location/Qualifiers
 1..5811
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /chromosome="3"
 /map="92E8-14"
 1..5811
 /gene="bon"
 1..693
 /gene="bon"
 694..4095
 /gene="bon"
 /note="TIF1-like protein; BON"
 /codon_start=1

/product="bonus"
 /protein_id="AAF19646.1"
 /db_xref="GI:6631000"
 /translation="MDMDLEQLKQKQADIPNPEIPRIQVQNVSDLVQVLSQLEIITVDGPRP
 PTPSPNGTPQHPERRPPSPMAPRLPGLPQMPAGLSPGPPVNGPQGPPLYS
 NAAQQFNNLSMSRSPGDSGKVRFGMPVPVGMQRHQPVHSSSTHQNMDISLRL
 LNNQASQPNAMAFNCPSPYDGGQAPAPAHQMPQMRPHFMPGQGGFSGGAGP
 GGPRAHPMANSNARFQSOYORANANAAOAAAAMAAGAAGGAGQIIPSGALORPOM
 SNPMONVAMONSIGFHSQAGFNTGPOTSPOLGGGGMSLAKHTIPQASQSMCS
 QGGLPLPPRANGQTSBNFKLSKSPNTLKNSTPBLGGCGAGHGGHNGSSAQLNMA
 ALGLPVAVSLSNVTSTLPTPSPSTHNTDFTPEIDKVRDSDINDLIATLAKDSN
 GVQVLPBRTKTTSPQVHSTDLSNTQVNNKNKQKDPNEDMCAVCLDGLKMDK
 CPKVFHONCHIPAISLPDESBSMQLCVNLIKELTGESKSSGSLALELILQ
 RICLELYOYEQSLNFRBPSAPNTSYEIVSSPMSLVIRLRDPSSPNHKDIAGF
 VSDVRLIFSNTYLPQEDTKYSNAKLENPREPOLATMLPQECTKXQKXNTNSP
 ALLGVNATGSPSPLENGKSCGSASLSDSDACLPKAKRRSAHE"
 931..1092
 /gene="bon"
 /note="Region: ring-finger domain"
 1225..1338
 /gene="bon"
 /note="Region: B-box 1"
 1411..1506
 /gene="bon"
 /note="Region: B-box 2"
 1510..1929
 /gene="bon"
 /note="Region: coiled-coil domain"
 3385..3528
 /gene="bon"
 /note="Region: PHD-finger domain"
 3571..3909
 /gene="bon"
 /note="Region: bromodomain"
 4096..5811
 /gene="bon"
 4543..4548
 /gene="bon"

ORIGIN

Query Match 3.7%; Score 56.4; DB 3; Length 5811;
 Best Local Similarity 59.3%; Pred. No. 0.6;
 Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 528 AATGAGGACGATGCGCGTGTGCGGAGCGCGGAGGACCTATCTGTGACGGCTGC 587
 DB 3379 AATGAGGACGATGCGCGTGTGTGATGAGAGGAGACTATGTGCGACAAGTGT 3438
 QY 588 CCTCGGCGCTTCACTGCGCTGCTGCTCCGCTCCGGAAGATCCCACTGAGGACC 647
 DB 3439 CCCAAGTTTTCACACAGAACTGTCAACATCCTTGAGATCAGCTCGTCCGAGCAGAGAC 3498
 QY 648 TGGAGTGTCTCAAGTCTGCGAGCGACACAGTCCAGAGGTG 689
 DB 3499 GAGAGCTGCGAGTGTCTACTGTGCGTCAACATCAAGAGACTG 3540

RESULT 95
 LOCUS CQ583074 6124 bp DNA linear PAT 02-FEB-2004
 DEFINITION Sequence 10832 from Patent WO0171042.
 ACCESSION CQ583074
 VERSION CQ583074.1 GI:41644029
 KEYWORDS
 SOURCE Drosophila sp.
 ORGANISM Drosophila sp.

	REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
	AUTHORS	Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
	TITLE	Detection kits, such as nucleic acid arrays, for detecting the expression of 10,000 or more Drosophila genes and uses thereof
	JOURNAL	Patent: WO 0171042-A 10832 27-SEP-2001;
	FEATURES	PE Corporation (NY) (US)
	SOURCE	Location/Qualifiers 1..6124 /organism="Drosophila sp." /mol_type="unassigned DNA" /db_xref="taxon:7242"
	ORIGIN	
	Query Match	3.7%; Score 56.4; DB 6; Length 6124;
	Best Local Similarity	59.3%; Pred.No.0.6;
	Matches	96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
Oy	528 AATGAGCAGAGTGGCCGTGTGTGGGACGCGGGAGCTCATCTGTGA CGGCTGC	587
Db	3391 AACGAGACTGTGGCGCGCTGTGTGTGATGAGCGCAGTCGATGCTCGACAAGTGT	3450
Oy	588 CCTGGGGCTTTCCACTTGCGCTGCTGCCCTTCGCTCCGAGATCCCAGTGGAGCC	647
Db	3451 CCCAAAGTTTTCCATCAAGACTGTCACTCCCTGGATCACTCGTTCCGAGCAGAGCC	3510
Oy	648 TGGAGGTGCTCCAGCTGCTCGACGAGCAACATCCAGGAGGTG	689
Db	3511 GAGAGCTGGCAGTGTCTA CTGTGCTCAACTCAAGAGCTG	3552
RESULT 96		
COS83073	LOCUS	COS83073 21747 bp DNA linear PAT 02-FEB-2004
COS83073	SEQUENCE	Sequence 10831 from Patent WOO171042.
ACCESSION		COS83073
VERSION		COS83073.1 GI:41644028
KEYWORDS		.
SOURCE		Drosophila sp.
ORGANISM		Drosophila sp. Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE		1 Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
AUTHORS		Detection kits, such as nucleic acid arrays, for detecting the expression of 10,000 or more Drosophila genes and uses thereof
TITLE		Patent: WO 0171042-A 10831 27-SEP-2001;
JOURNAL		PE Corporation (NY) (US)
FEATURES		Location/Qualifiers 1..21747 /organism="Drosophila sp." /mol_type="unassigned DNA" /db_xref="taxon:7242"
SOURCE		
ORIGIN		
Query Match		3.7%; Score 56.4; DB 6; Length 21747;
Best Local Similarity		59.3%; Pred.No.0.49;
Matches		96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
Oy	528 AATGAGCAGAGTGGCCGTGTGTGGGACGCGGGAGCTCATCTGTGA CGGCTGC	587
Db	17810 AACGAGACTGTGGCGCGCTGTGTGTGATGAGCGCAGTCGATGCTCGACAAGTGT	17866
Oy	588 CCTGGGGCTTTCCACTTGCGCTGCTGCCCTTCGCTCCGAGATCCCAGTGGAGCC	647
Db	17870 CCCAAAGTTTTCCATCAAGACTGTCACTCCCTGGATCACTCGTTCCGAGCAGAGCC	17922
Oy	648 TGGAGGTGCTCCAGTGTGCTCGACGAGCAACATCCAGGAGGTG	689
Db	17930 GAGAGCTGGCAGTGTCTA CTGTGCTCAACTCAAGAGCTG	17971

RESULT 97
AC017943 97538 bp DNA linear HTG 09-DEC-1999

LOCUS AC017943
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC017943
VERSION AC017943.1 GI:6553247
KEYWORDS HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 97538)
Adams,M. and Venter,J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212734 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Source
1..97538
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match 3.7%; Score 56.4; DB 2; Length 97538;
Best Local Similarity 59.3%; Pred. No. 0.38;
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Oy 528 AATGAGACGAGTGTGCGGTGTGTGCGGACGCGCGGAGCTCATCTGTGACGCGTGC 587
|||
Db 9373 AACGAGACGTGTGCGCGGTGTGTGCGGACGCGCGGAGCTCATCTGTGACGCGTGC 9422
588 CTTGCGGCGCTTCCACCTGCGCTGTGCGCGCGCGCGCGGAGATCCCACTGGAGCC 647
9433 CCCAAAGTTTTCATCATGACTGCATCCCTGCGATCAGCTGTGCGCGAGAGAGC 94922
Oy 648 TGGAGGTCTTCACCTGCGCTGTGCGGACGACAGTCCAGAGGTG 689
|||
Db 9493 GAGAGCTGGCAGTGTCTTGTGTGCTGCATCATGAGAGGCTG 9534

RESULT 98
AC007757 170939 bp DNA linear INV 15-FEB-2001

LOCUS AC007757
DEFINITION Drosophila melanogaster, chromosome 3R, region 92E-92F, BAC clone
BACR13J17, complete sequence.
ACCESSION AC007757
VERSION AC007757.5 GI:12831350
KEYWORDS HTG.
ORGANISM Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 170939)
Ceiniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Goecky,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,U., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champ,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dyesnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howard,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mettel,B., Moehref,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Munoo,J.,
Pachle,J., Paragae,V., Park,S., Patel,S., Pfeiffer,B.,

Phanenaevov, S., Piltman, G. S., Puri, V., Richards, S., Scheeler, F.,
Shaplenon M., Strong, R., Switska, R., Tector, C., Williams, S. M.,
Zaveri, J. S., Smith, H. O., Rubin, G. M. and Venter, U. C.
Sequencing of Drosophila chromosome 3R, region 92E-92F
Unpublished
2 (bases 1 to 170939)
Celniker, S. E., Agbayan, A., Arcineta, T. T., Baxter, E., Blazew, R. G.,
AUTHORS

VERSION	AE003731.3	GI:23175594
KEYWORDS	Drosophila melanogaster (fruit fly)	
SOURCE	Drosophila melanogaster	
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 230001)	
AUTHORS	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., Davis,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Morten,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazer,R.G., Chape,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabo,R.G.L., Abille,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Bakendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolintineanu,S., Borkova,D., Botchan,M.R., Boulter,J., Brokstein,P., Brotter,P., Burks,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,J.B., Davies,P., de Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Diez,S.M., Dodson,K., Doup,L.E., Domingue,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriere,S., Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kenton,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milburn,N.V., Mockay,C., Mout,R., Mostoslavsky,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nuskern,D.R., Paclob,J.M., Palazzolo,M., Peltman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,P., Shen,H., Shue,B.C., Sidon-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,E., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,R., Wang,A.H., Wang,X., Wang,Z.Y., Wasarman,D.A., Weinstock,G.M., Weisenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.	
TITLE	The genome sequence of Drosophila melanogaster	
JOURNAL	Science 287 (5461), 2185-2195 (2000)	
MEDLINE	20196006	
PUBMED	10731132	
REFERENCE	2 (bases 1 to 230001)	
AUTHORS	Celniker,S.E., Wheeler,D.A., Kronmiller,B., Carlson,J.W., Halpern,A., Patel,S., Adams,M., Champ,M., Dugan,S.P., Frisoe,E., Hodgson,A., George,R.A., Hoskins,R.A., Laverly,T., Muzny,D.M., Nelson,C.R., Paclob,J.M., Park,S., Pfeiffer,B.D., Richards,S., Sodergren,E.J., Svirskas,R., Taboy,P.B., Wan,K., Stapleton,M., Sutton,G.G., Venter,C., Weinstock,G., Scherer,S.E., Myers,E.W., Gibbs,R.A. and Rubin,G.M.	
TITLE	Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence	
JOURNAL	Genome Biol. 3 (12), RESEARCH0079 (2002)	
MEDLINE	22426065	
PUBMED	12537568	
REFERENCE	3 (bases 1 to 230001)	
AUTHORS	Mista,S., Crosby,M.A., Mungall,C.J., Matthews,B.B., Campbell,K.S., Hradecky,P., Huang,Y., Kaminker,J.S., Millburn,G.H., Prochnik,S.E., Smith,C.D., Tupy,J.L., Whitfield,E.J., Bayraktaroglu,L., Berman,B.P., Bellocourt,B.R., Celniker,S.E., de Grey,A.D., Drysdale,R.A., Harris,N.L., Richter,J., Russo,S., Schroeder,A.J., Shu,S.C., Stapleton,M., Yamada,C., Ashburner,M., Gelbart,M.M., Rubin,G.M. and Lewis,S.E.	
TITLE	Annotation of the Drosophila melanogaster euchromatic genome: a systematic review	
JOURNAL	Genome Biol. 3 (12), RESEARCH0083 (2002)	
MEDLINE	22426069	
PUBMED	12537572	
REFERENCE	4 (bases 1 to 230001)	
AUTHORS	Kaminker,J.S., Bergman,C.M., Kronmiller,B., Carlson,J., Svirskas,R., Patel,S., Frisoe,E., Wheeler,D.A., Lewis,S.E., Rubin,G.M., Ashburner,M. and Celniker,S.E.	
TITLE	The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective	
JOURNAL	Genome Biol. 3 (12), RESEARCH0084 (2002)	
MEDLINE	22426070	
PUBMED	12537573	
REFERENCE	5 (bases 1 to 230001)	
AUTHORS	Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-MAR-2000) Celera Genomics, 45 West Gate Drive, Rockville, MD 20850, USA	
REFERENCE	6 (bases 1 to 230001)	
AUTHORS	FlyBase	
CONSTRM	Direct Submission	
TITLE	Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA	
JOURNAL	7 (bases 1 to 230001)	
REFERENCE	FlyBase	
CONSTRM	Direct Submission	
TITLE	Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA	
JOURNAL	On Sep 18, 2002 this sequence version replaced gi:10726652.	
COMMENT	Location/Qualifiers	
FEATURES	1..230001	
SOURCE	/organism="Drosophila melanogaster"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:7227"	
	/chromosome="3R"	
	11240..11520	
	/gene="MtnD"	
	/locus_tag="CG33192"	
	/db_xref="FBgn0053192"	
	join(11240..11276,11343..11520)	
	/gene="MtnD"	
	/locus_tag="CG33192"	
	/product="CG33192-RA"	
	/db_xref="FBgn0053192"	
	join(11252..11276,11343..11433)	
	/gene="MtnD"	
	/locus_tag="CG33192"	
	/codon_start=1	
	/product="CG33192-PA"	
	/protein_id="AA041579.1"	
	/db_xref="GI:20381377"	
	/db_xref="FBgn0053192"	
	/translation="MGCKACGTCGCSATKCGDNCAGSQCCGSC"	
	complement(11918..28547)	
	/gene="Stat92E"	
	/locus_tag="CG4257"	
	/map="92F1..92F1"	
	/db_xref="FBgn0016917"	
	complement(join(11918..12760,13429..13698,15456..15570,15635..15835,15894..16151,16210..16689,16758..17788,17915..18327,28316..28547))	
	/gene="Stat92E"	
	/locus_tag="CG4257"	
	/product="CG4257-RE"	
	/note="This object was named CG4257-RD in FlyBase d.melanogaster release 3.1"	
	/db_xref="FBgn0016917"	
	complement(join(11918..12760,13429..13698,15456..15591,15635..15835,15894..16151,16210..16689,16758..17788,17915..18327,28316..28547))	
	/gene="Stat92E"	
	/locus_tag="CG4257"	

```
/product="CG4257-RP"
/db_xref="FLYBASE:FBgn0016917"
complement(join(11918..12780,13429..13698,15456..15570,
15635..15835,15894..16151,16210..16689,16758..17788,
26956..27365))
/feature="Stat92E"
/locus_tag="CG4257"
/product="CG4257-RB"
/db_xref="FLYBASE:FBgn0016917"
complement(join(11918..12780,13429..13698,15456..15570,
15635..15835,15894..16151,16210..16689,16758..17788,
17915..18327,26956..27365))
/feature="Stat92E"
/locus_tag="CG4257"
/product="CG4257-RC"
/db_xref="FLYBASE:FBgn0016917"
complement(join(11644..13698,15456..15570,15635..15835,
15894..16151,16210..16689,16758..17788,17915..18039))
/feature="Stat92E"
/locus_tag="CG4257"
/codon_start=1
/product="CG4257-PC"
/protein_id="AANI4350.2"
/db_xref="GI:28381378"
/translation="MSLMKRISSHYDCQRMAYYEKGMLERLCLAPWIDRIWSE
OITPNTDLEEVALKENEDLOKILSTRASDQALKPRVIELCALIORISAVELYTH
LSGLOKEIQLTERKSVANTAGQSMPLNPNMANTPMVGTGYMDPSDLAVNSCPP
VVOGIGPIHNVQGTIASPALGNVTPKVELYEQIOMOSLINEFGCANALKILQNY
SYMNSTSPNNAEAARSLIDEKAAVLTVLRSFMYEYLHEVILHELNRHQAOA
GNGAPNEGLDDIORCFEMLESFIAMHLAAVELMRVRLVTEPELTHLEOVQNAO
KULVGSATVDKOPQOVMTKNTFRFASVRMLGSGLGHNHNPTEVCIIIMSEIOSOF
VIRNTOMNNSLSGSGSERIQNASSTMEOQNHNVSASFRNMOLKKIKRAEKGES
VNDKEXALPPTTTTNDPOIRVWTLSLPVVIVHGNQEPQSWATTMDNAPAEIYRD
PNTIDRVTAQLSVALNIRFSGCTGRSLTIDNDFLYEKLQKEERSEYITWNOFCKE
PMDRSFTPEWEPFALIMKLTQKMLGMWKGCGIMGFINTKTAQTDLRSVYIGIFLL
RFSDSLGGVTLIAYVNIENGLVTMLAPMTARDFVNLADRIDLDVLCWLMHPSDRNAS
PYKRDVAFGEFFYSKRODPVTGYKSTLHVHCNENGSTSGTPHHAQESMQLNGNDF
GMADPDTITNPFNF"
complement(join(113644..13698,15456..15570,15635..15835,
15894..16151,16210..16689,16758..17788,17915..18039))
/locus_tag="CG4257"
/feature="Stat92E"
/locus_tag="CG4257"
/notes="This object was named CG4257-PD in FlyBase
D.melanogaster release 3.1"
/codon_start=1
/product="CG4257-PE"
/protein_id="AAFS5773.2"
/db_xref="GI:28381379"
/db_xref="FLYBASE:FBgn0016917"

Query Match 3.7%; Score 56.4; DB 3; Length 230001;
Best Local Similarity 59.3%; Pred. No. 0.33; Mismatches 66; Indels 0; Gaps 0;
Matches 96; Conservative 0;

QY 528 AATGAGGACGAGTGTCCGATGTCCGAGCGGCGGAGGCTCATCTGCTGTGCGGCTGC 587
Db 86207 AACGAGGACTGTCGCGCCGCTCTCTCGATGAGGCGAGCTGATGTGCTGCGACAACTGT 86266
QY 588 CCTGGGCGCTTTCACCTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGGACC 647
Db 86267 CCCAAGTTTTCATCAGAACTGTCAATCCCTCGATCAGCTCGTTGCCGAGCGAGAGC 86326
QY 648 TGGAGGTGCTCCGCTGCTGCGAGCAAGTCCAGAGGTG 689
Db 86327 GAGAGCTGCGAGTCTACTGTGCTGCTCAATCAAGAGACTG 86368
```

Search completed: February 9, 2005, 16:24:26
Job time : 6517 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 09:10:56 ; Search time 739 Seconds

(without alignments)
10974.771 Million cell updates/sec

Title: US-09-508-658a-3

Perfect score: 1545
Sequence: 1 agaggaagcgcaggtcttcctc.....aaaataataaacatgcgcg 1545

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 200 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1541.8	99.8	1545	2	AAX26937
2	1367.8	88.5	1463	2	AAX26937 cDNA enco
3	1023.6	66.3	2036	2	AAX26936 cDNA enco
4	1023.6	66.3	2245	2	AAX26936 cDNA enco
5	385.6	25.0	1656	2	AAX58605 Human aut
6	118.4	7.7	9876	6	AAX58605 Mouse aut
7	101.8	6.6	9876	6	ABL32092 Human imm
8	4.2	9646	12	ADH12179	ADH12179 Human CHD
9	3.9	60	6	ABN38615	ABN38615 Human spl
10	3.9	5739	2	ABK70271	ABK70271 Human lun
11	3.9	6327	2	AAI32301	AAI32301 Dermatomy
12	3.9	6417	6	ABL64410	ABL64410 Stomach c
13	3.9	6417	6	ABX04169	ABX04169 Human mRN
14	3.9	6417	6	ABK84696	ABK84696 Human chrN
15	3.9	6417	10	AAAD54631	AAAD54631 Human chr
16	3.9	6417	10	AAAD54632	AAAD54632 Human chr
17	3.9	6417	12	ADP10531	ADP10531 Reference
18	3.9	6475	4	AAI58444	AAI58444 Human pol
19	3.9	6475	4	ADQ98655	ADQ98655 DNA enco
20	3.9	6475	9	ADBA8415	ADBA8415 Novel hum
21	3.9	6553	4	AAI60230	AAI60230 Human pol

95	47	3.0	6610	4	AA829013	AA829013 Genomic s
96	47	3.0	6610	4	AA826967	AA826967 Human gen
97	47	3.0	6610	4	AA826968	AA826968 Human gen
98	47	3.0	6610	4	ABA06837	ABA06837 Human gen
99	47	3.0	6610	4	ABA06838	ABA06838 Human gen
100	47	3.0	6610	6	ABT07848	ABT07848 Novel hum
101	47	3.0	6610	6	ABT07849	ABT07849 Novel hum
102	47	3.0	6610	6	ABV84174	ABV84174 Human pol
103	47	3.0	6610	6	ABV84175	ABV84175 Human pol
104	46.6	3.0	2673	4	ABL29909	ABL29909 Drosophila
105	46.6	3.0	4705	4	ABL29908	ABL29908 Drosophila
106	46.2	3.0	550	12	ACH80015	ACH80015 Human gen
107	46	3.0	795	8	ACC6331	ACC6331 Human dlt
108	45.8	3.0	10732	3	AAA10594	AAA10594 Gene enco
109	45.4	2.9	5261	10	ADD01271	ADD01271 Human nuc
110	45	2.9	5476	2	AA760078	AA760078 Human SMC
111	45	2.9	5476	10	ADD15195	ADD15195 Human SMC
112	45	2.9	5476	12	ADM33396	ADM33396 Human PRO
113	45	2.9	5476	12	ADP10502	ADP10502 Reference
114	44.6	2.9	2165	12	ADP07311	ADP07311 Human HIC
115	44.6	2.9	2327	12	ADL13335	ADL13335 Human etc
116	44.6	2.9	50295	9	ADH02993	ADH02993 Mouse Sep
117	44.6	2.9	50295	10	ADB72731	ADB72731 Mouse Sep
118	44.6	2.9	50295	10	ADC85473	ADC85473 Mouse Sep
119	44.6	2.9	50295	12	ADM74588	ADM74588 Murine ca
120	44.2	2.9	103	12	ACH93715	ACH93715 Human gen
121	44.2	2.9	510	6	ABO75937	ABO75937 CHD activ
122	44	2.8	328	11	ADM65053	ADM65053 NRY polym
123	44	2.8	358	5	AA575441	AA575441 DNA encod
124	44	2.8	387	9	ACH36012	ACH36012 Human end
125	43.8	2.8	1589	11	ADM01317	ADM01317 Human cDN
126	43.8	2.8	5911	6	ABL64387	ABL64387 Stomach c
127	43.8	2.8	5911	6	ABK64506	ABK64506 Human ben
128	43.8	2.8	18876	10	AD123928	AD123928 Streptomy
129	43.8	2.8	61944	10	AD123920	AD123920 Streptomy
130	43.6	2.8	83709	12	ADQ19964	ADQ19964 Human sof
131	43.6	2.8	6687	12	ADN73398	ADN73398 Thale cre
132	43.6	2.8	6687	12	ADN73434	ADN73434 Thale cre
133	43.6	2.8	15738	10	AD123898	AD123898 Streptomy
134	43.6	2.8	37360	10	AD123892	AD123892 Streptomy
135	43.4	2.8	435	8	ABX49595	ABX49595 Bovine ES
136	43.4	2.8	1283	4	AAH17044	AAH17044 Human cDN
137	43.4	2.8	7382	4	ABL04467	ABL04467 Drosophila
138	43.4	2.8	8035	8	ACC81113	ACC81113 Human RER
139	43.4	2.8	10165	4	ABL04466	ABL04466 Drosophila
140	43.2	2.8	989	3	AAA02539	AAA02539 Human col
141	43	2.8	327	11	ADM65056	ADM65056 NRY polym
142	43	2.8	1421	12	ADP13324	ADP13324 Renal cel
143	43	2.8	1825	6	ABQ54550	ABQ54550 Human ova
144	43	2.8	2433	6	ABL39701	ABL39701 Human NS
145	43	2.8	2500	10	ADB87421	ADB87421 Transgene
146	43	2.8	2500	12	AD135230	AD135230 Human eta
147	43	2.8	2500	12	AD196848	AD196848 Stablinzin
148	43	2.8	2664	3	AACT6332	AACT6332 Human ORF
149	43	2.8	5319	10	ADC51537	ADC51537 Human Neu
150	43	2.8	5400	10	ADC51541	ADC51541 Human Neu
151	42.8	2.8	282	6	ABKS3809	ABKS3809 Human hea
152	42.8	2.8	511	6	ABKS4025	ABKS4025 Human hea
153	42.8	2.8	1365	8	ABT32152	ABT32152 Benzodiaz
154	42.8	2.8	1611	12	ADN99091	ADN99091 Novel hum
155	42.8	2.8	1611	12	ADN00660	ADN00660 Novel hum
156	42.8	2.8	1653	4	AAKS2490	AAKS2490 Human pol
157	42.8	2.8	1691	4	AAKS1506	AAKS1506 Human pol
158	42.8	2.8	2374	10	ADB62956	ADB62956 Human cDN
159	42.8	2.8	6225	2	AAKS5273	AAKS5273 Human enz
160	42.8	2.8	6225	3	AAK34721	AAK34721 Human ade
161	42.8	2.8	6225	3	AAK20843	AAK20843 Human mu1
162	42.8	2.8	32539	8	ABT96537	ABT96537 Human nuc
163	42.8	2.8	32539	8	ABT32129	ABT32129 Benzodiaz
164	42.8	2.8	35384	3	AAFT1436	AAFT1436 Human enz
165	42.8	2.8	35459	10	ABE97130	ABE97130 Human nuc
166	42.6	2.8	882	2	AAV40176	AAV40176 Flavobact
167	42.6	2.8	1182	3	AAA75908	AAA75908 CDNA enco

168	42.6	2.8	1469	2	AAQ40323	AAQ40323 Sequence
169	42.4	2.7	1581	10	AD160528	AD160528 Secreted
170	42.4	2.7	16418	4	AAK81568	AAK81568 Human imm
171	42.4	2.7	347001	12	ADP43517	ADP43517 Human MAD
172	42.2	2.7	2137	8	ACA40697	ACA40697 Prokaryot
173	42.2	2.7	3137	12	ADQ24019	ADQ24019 Human sof
174	42.2	2.7	3465	12	ADP28768	ADP28768 Human sec
175	42.2	2.7	4523	10	ADP60227	ADP60227 Human gen
176	42.2	2.7	4523	10	ADP60229	ADP60229 Human gen
177	42.2	2.7	4523	10	ADP60231	ADP60231 Human gen
178	42.2	2.7	4523	10	ADP60233	ADP60233 Human gen
179	42.2	2.7	4523	10	ADP45822	ADP45822 Human gen
180	42.2	2.7	6749	12	ADQ24048	ADQ24048 Human sof
181	42.2	2.7	9314	4	ABL30497	ABL30497 Drosophila
182	42.2	2.7	14509	4	ABL30496	ABL30496 Drosophila
183	42.2	2.7	110000	4	AA199682	AA199682, 29
184	42.2	2.7	110000	4	AA199683	AA199683, 29
185	42	2.7	3612	3	AA251921	AA251921 N. tabaco
186	41.8	2.7	4598	2	AAK56436	AAK56436 Human atr
187	41.6	2.7	1941	2	AAQ24522	AAQ24522 CDNA enco
188	41.6	2.7	4269	5	AAK66532	AAK66532 DNA encod
189	41.4	2.7	615	8	ABE22262	ABE22262 Schwann c
190	41.4	2.7	997	8	ABE22275	ABE22275 Mouse and
191	41.4	2.7	1173	5	AAK87672	AAK87672 DNA encod
192	41.4	2.7	4616	2	AAAT29604	AAAT29604 HIC-1 gen
193	41.4	2.7	14835	6	AAK94858	AAK94858 Human DNA
194	41.2	2.7	7788	10	AD123894	AD123894 Streptomy
195	41.2	2.7	14707	6	AAK53529	AAK53529 Genomic D
196	41.2	2.7	16132	9	AAK58467	AAK58467 Mouse per
197	41.2	2.7	37360	10	AD123892	AD123892 Streptomy
198	41.2	2.7	58857	3	AAK58471	AAK58471 Nucleotid
199	41	2.7	536	10	ADB68842	ADB68842 Minority
200	41	2.7	110000	4	AA199682_12	Continuation (13 o

ALIGNMENTS

RESULT 1	
AAK26937	
ID	AAK26937 etandard; DNA; 1545 BP.
AAK26937;	
24-JUN-1999	(first entry)
CDNA	encoding a human autoimmune regulator-2 (AIR-2) protein.
Autoimmune regulator-2; AIR-2; immune maturation; immune response;	
disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;	
APECED; autoimmune polyglandular syndrome type I; APS I; ss.	
Homo sapiens.	
Key	Location/Qualifiers
CDS	237..1283
FT	/*tag= a
FT	/product= "AIR-2"
PN	W0915559-A1.
PD	01-APR-1999.
PR	23-SEP-1997; 97FI-00003762.
PA	(FIM-) FINNISH IMMUNOTECHNOLOGY LTD.
PI	Krohn K., Heino M., Peterson P., Scott H., Antonarakis S., Lalioti M;
PI	Shimizu N., Kudoh J;
DR	WPI; 1999-244390/20.

DR P-PSDB; AA01713.

Autoimmune regulator 1 (AIR1) DNA sequence.

Claim 4; Page 27-29; 59pp; English.

The present sequence encodes an autoimmune regulator-2 (AIR-2) protein. The AIR polypeptides and polynucleotides can be used in methods for the diagnosis and treatment of diseases related to immune maturation and regulation of immune response towards self and nonself. They can be used particularly in the diagnosis and treatment of autoimmune disease CC polyendocrinopathy candidiasis ectodermal dysplasia (APECED) (also known as autoimmune polyglandular syndrome type I (APS I)).

XX Sequence 1545 BP; 261 A; 485 C; 508 G; 291 T; 0 U; 0 Other;

Query Match 99.8%; Score 1541.8; DB 2; Length 1545;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1543; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 AAGAAAGTGAAGCTTTCTCAGGCTTTAAAGCATGGCTTTGGTTCAGGCTGTAACCG 60
DB 1 AAGAAAGTGAAGCTTTCTCAGGCTTTAAAGCATGGCTTTGGTTCAGGCTGTAACCG 60
QY 61 CTGCTCTCAGCTGGGCGCGGGTGGGCGCGGGCGCGCGCGCTGCTATAGCCAGAGGTCAAG 120
DB 61 CTGCTCTCAGCTGGGCGCGGGTGGGCGCGGGCGCGCGCGCTGCTATAGCCAGAGGTCAAG 120
QY 121 ATCCATGGGAAATGCATGCTCATCTTTGCTCCCAAGCATGGTTCTTAATGGGGTAGAA 180
DB 121 ATCCATGGGAAATGCATGCTCATCTTTGCTCCCAAGCATGGTTCTTAATGGGGTAGAA 180
QY 181 GCAAGTGGGAGAGACCTCTGCTGGGCTGGCCCACTGCTGTAGAGAAAGGTTAATGT 240
DB 181 GCAAGTGGGAGAGACCTCTGCTGGGCTGGCCCACTGCTGTAGAGAAAGGTTAATGT 240
QY 241 GGTGGTGTACAGTTCTGGGGGCGCTGGAACGAGACGCTGCAAGAAACGGGTTTCT 300
DB 241 GGTGGTGTACAGTTCTGGGGGCGCTGGAACGAGACGCTGCAAGAAACGGGTTTCT 300
QY 301 TCCCAATAGGATGAGCCCGCGGGGGTGTCTGTGAGAACAGATGATGGGAAACAGGTG 360
DB 301 TCCCAATAGGATGAGCCCGCGGGGGTGTCTGTGAGAACAGATGATGGGAAACAGGTG 360
QY 361 GTGAGGGCAGAAATTTGAGGCGCTGAGCAATGGGACAGAGGCGAGAGCTGGGAGTTCA 420
DB 361 GTGAGGGCAGAAATTTGAGGCGCTGAGCAATGGGACAGAGGCGAGAGCTGGGAGTTCA 420
QY 421 GTACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAAAGAGTGCTCTCAGAGAGGCTG 480
DB 421 GTACCCAGAGATGCTGCTGGGGAGCTGTTTGGGAAAGAGTGCTCTCAGAGAGGCTG 480
QY 481 TGCACCCAGCCAGCTGTCATGAGGAGTCTCTGCTGTCAGAGAAATGAGAGAGT 540
DB 481 TGCACCCAGCCAGCTGTCATGAGGAGTCTCTGCTGTCAGAGAAATGAGAGAGT 540
QY 541 GTGCGCTGTGTGGGAGCGCGGGGAGCTCATCTGTGTGACGCGTGGCTCGGCTTTC 600
DB 541 GTGCGCTGTGTGGGAGCGCGGGGAGCTCATCTGTGTGACGCGTGGCTCGGCTTTC 600
QY 601 ACCCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 601 ACCCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
QY 661 GCTGCTGAGGAGCAAGTTCAGAGAGTGCAGCCCGGGCAGAGAGCCCGCCAG 720
DB 661 GCTGCTGAGGAGCAAGTTCAGAGAGTGCAGCCCGGGCAGAGAGCCCGCCAG 720
QY 721 AGCCACCTGAGAGACCCCGCTCCCGGGGGTGTAGTGTGGGGAGAGAGTAAAG 780
DB 721 AGCCACCTGAGAGACCCCGCTCCCGGGGGTGTAGTGTGGGGAGAGAGTAAAG 780
QY 781 GTCCACTGGGAAACCCCTAGCCGAGTGAAGAGACTTGTCTAAGCAAGCACTCGCG 840

```

```

DB 781 GTCCACTGGGAAACCCCTAGCCGAGTGAAGAGACTTGTCTAAGCAAGCACTCGCG 840
QY 841 CTCGCCCTTCTGAGCCCGCTGCGAGTCTGAGTCTCTGAGCCCTGACCCCTACTGT 900
DB 841 CTCGCCCTTCTGAGCCCGCTGCGAGTCTGAGTCTCTGAGCCCTGACCCCTACTGT 900
QY 901 GTGTGGTCTTGAAGGTCAGAGAACTGCTCTGTGTGCGGTTGCGGGGTGTGAGAG 960
DB 901 GTGTGGTCTTGAAGGTCAGAGAACTGCTCTGTGTGCGGTTGCGGGGTGTGAGAG 960
QY 961 ATGTGAGACGCTGCGGCTGAGTACCTGAGCGCGCTGCTTCCAGTGGCTGACAT 1020
DB 961 ATGTGAGACGCTGCGGCTGAGTACCTGAGCGCGCTGCTTCCAGTGGCTGACAT 1020
QY 1021 TCCACCGGACCTCCCGCGCGAGCGAGCTGCGATGCAATCTGCTCAGAGAGC 1080
DB 1021 TCCACCGGACCTCCCGCGCGAGCGAGCTGCGATGCAATCTGCTCAGAGAGC 1080
QY 1081 TGACCCAGCCCTGTGAGAGGGGTCTGAGCCCGCCGCGCTGCGCCCTGGG 1140
DB 1081 TGACCCAGCCCTGTGAGAGGGGTCTGAGCCCGCCGCGCTGCGCCCTGGG 1140
QY 1141 CTGCCAAGGATGACATGACAGTCAAGAGCGCGCTGCAAGAGATGACCTGAGTCC 1200
DB 1141 CTGCCAAGGATGACATGACAGTCAAGAGCGCGCTGCAAGAGATGACCTGAGTCC 1200
QY 1201 TTCTGAGCAGACACCTTGTGATGAGCATCTGAGTGGCCATCAGAGATGAGCCGTG 1260
DB 1201 TTCTGAGCAGACACCTTGTGATGAGCATCTGAGTGGCCATCAGAGATGAGCCGTG 1260
QY 1261 CGGCGGCGCGCTTCCCTCTGAGCCAGATGGCGGGACATGACACTGTGAGAGAG 1320
DB 1261 CGGCGGCGCGCTTCCCTCTGAGCCAGATGGCGGGACATGACACTGTGAGAGAG 1320
QY 1321 TGTGAGAGAGACACCTCTCTCTGAGTCTGAGTCTGAGAGCCGCGCTGAGTCAAGAGG 1380
DB 1321 TGTGAGAGAGACACCTCTCTCTGAGTCTGAGTCTGAGAGCCGCGCTGAGTCAAGAGG 1380
QY 1381 GACAGCGCACCTTGTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 1440
DB 1381 GACAGCGCACCTTGTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 1440
QY 1441 GGCATATGTGCTGGAATTAACCTGCGCCCACTTCTCTACTCTGGAATCCCGGAG 1500
DB 1441 GGCATATGTGCTGGAATTAACCTGCGCCCACTTCTCTACTCTGGAATCCCGGAG 1500
QY 1501 GCCTCTCTGCTGCTGAGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 1545
DB 1501 GCCTCTCTGCTGCTGAGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 1545

RESULT 2
AAK26938
ID AAK26938 standard; DNA; 1463 BP.
AC AAK26938;
XX
XX 24-JUN-1999 (first entry)
XX
DE cDNA encoding a human autoimmune regulator-3 (AIR-3) protein.
XX
XX Autoimmune regulator-3; AIR-3; immune maturation; immune response;
XX disease; autoimmune polyendocrinopathy candidiasis ectodermal dysplasia;
XX APECED; autoimmune polyglandular syndrome type I; APS I; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 237..1001
XX FT /*tag= a
XX FT /product= "AIR-3"
XX

```

PN WO9915559-A1.
XX 01-APR-1999.
PD
XX 23-SEP-1998; 98WO-FI000749.
XX 23-SEP-1997; 97FI-00003762.
PR
XX (FIRM-) FINNISH IMMUNOTECHNOLOGY LTD.
XX Krohn K, Heino M, Peterson P, Scott H, Antonarakis S, Lallioi M,
PI Shimizu N, Kudoh J;
XX MPI; 1999-244390/20.
DR P-PSDB; AAY01714.
XX
PT Autoimmune regulator 1 (AIR1) DNA sequence.
XX
PS Claim 4; Page 31-32; 59pd; English.
XX
CC The present sequence encodes an autoimmune regulator-3 (AIR-3) protein.
CC The AIR polypeptides and polynucleotides can be used in methods for the
CC diagnosis and treatment of diseases related to immune maturation and
CC regulation of immune response towards self and nonself. They can be used
CC particularly in the diagnosis and treatment of autoimmune
CC polyendocrinopathy candidiasis ectodermal dysplasia (APECED) (also known
CC as autoimmune polyglandular syndrome type I (APS I))
XX
SQ Sequence 1463 BP; 245 A; 456 C; 486 G; 276 T; 0 U; 0 Other:

Query Match 88.5%; Score 1367.8; DB 2; Length 1463;
Best Local Similarity 94.6%; Pred. No. 4.1e-282;
Matches 1461; Conservative 0; Mismatches 2; Indels 82; Gaps 1;

QY 1 AAGAGAAAGGAGGCTCTTCTCAGGCTCTTAAGAGCATGGGCTTGTCTCAGGCTGTACCG 60
DB 1 AAGAGAAAGGAGGCTCTTCTCAGGCTCTTAAGAGCATGGGCTTGTCTCAGGCTGTACCG 60
QY 61 CTGCTCTCAGCTGGGCGCCGTGGGTGGGCGGCGGCCCTGCTATAGCCAGAGGTCAAG 120
DB 61 CTGCTCTCAGCTGGGCGCCGTGGGTGGGCGGCGGCCCTGCTATAGCCAGAGGTCAAG 120
QY 121 ATCCACTGGGAATGCCATGCTATCTTTCGCCAGCATGGTTCTTAAATGGGGTAGAA 180
DB 121 ATCCACTGGGAATGCCATGCTATCTTTCGCCAGCATGGTTCTTAAATGGGGTAGAA 180
QY 181 GCAGGTCGGGAGAGACTCTCTGGGCTGGCCCTGACCTGCTGTAGAGAAAGGGTTCATGT 240
DB 181 GCAGGTCGGGAGAGACTCTCTGGGCTGGCCCTGACCTGCTGTAGAGAAAGGGTTCATGT 240
QY 241 GGTGGGTGTACAGTTCCGGGGCCCTGGAACGAGCAGCTGTGAAGAAACCGGGTTTCT 300
DB 241 GGTGGGTGTACAGTTCCGGGGCCCTGGAACGAGCAGCTGTGAAGAAACCGGGTTTCT 300
QY 301 TCCCAATAGGGATGGCCCGGGGGGTGTCTGTTGAGACCAATGATGGGAAACAGGTG 360
DB 301 TCCCAATAGGGATGGCCCGGGGGGTGTCTGTTGAGACCAATGATGGGAAACAGGTG 360
QY 361 GTCAAGGGCAGAAATTTAGAGCCCTGAGCAATGGAGACAGAGGCGAGATCGGGAGTTCA 420
DB 361 GTCAAGGGCAGAAATTTAGAGCCCTGAGCAATGGAGACAGAGGCGAGATCGGGAGTTCA 420
QY 421 GTACCCAGAGATGCTCTGGGGAGGCTGTTTGGAAAGAGGTGGCTCTCAGAGAGGTGC 480
DB 421 GTACCCAGAGATGCTCTGGGGAGGCTGTTTGGAAAGAGGTGGCTCTCAGAGAGGTGC 480
QY 481 TGCACCCAGCCAGCTCTGATGGGCGTCTTTGCTGTGCCAGAAATGAGACAGAT 540
DB 481 TGCACCCAGCCAGCTCTGATGGGCGTCTTTGCTGTGCCAGAAATGAGACAGAT 540
QY 541 GTGCCGTGTGTGGGACGGCGGGAGACTCATGTGCTGACGGCTGCCCTCGGGCTTCC 600
DB 541 GTGCCGTGTGTGGGACGGCGGGAGACTCATGTGCTGACGGCTGCCCTCGGGCTTCC 600

QY 601 ACTGGGCGTGGCTGTCCCTCCGCTCCGGAGATCCCAAGTGGGACCTGGAGGTGCTCCA 660
DB 601 ACTGGGCGTGGCTGTCCCTCCGCTCCGGAGATCCCAAGTGGGACCTGGAGGTGCTCCA 660
QY 661 GCTGCTGAGGCAACAGTCCAGAGGTGCAAGCCCGGAGAGAGAGCCCGGCTCCAG 720
DB 661 GCTGCTGAGGCAACAGTCCAGAGGTGCAAGCCCGGAGAGAGAGCCCGGCTCCAG 720
QY 721 AGCCACCCGTGGAGACCCCGCTCCCGGGGCTTATGCTCGCGGAGAGAGGTAAAG 780
DB 721 AGCCACCCGTGGAGACCCCGCTCCCGGGGCTTATGCTCGCGGAGAGAGGTAAAG 780
QY 781 GTCCACTGGGGAACCCCTAGCCGAGATGACACGACTCTTGTCTCAAGACCTGCCG 840
DB 781 GTCCACTGGGGAACCCCTAGCCGAGATGACACGACTCTTGTCTCAAGACCTGCCG 840
QY 841 CTCCGCTTCTGACGCCCGCTGCCAGGTCTGAGACTCTCGGCTTGACCCCTACTGT 900
DB 841 CTCCGCTTCTGACGCCCGCTGCCAGGTCTGAGACTCTCGGCTTGACCCCTACTGT 900
QY 901 GTGTGGTCTTGAAGGTTCAGAGAACTTGGCTTCTGTGCGGTGCGGGGTGTGCGAG 960
DB 901 GTGTGGTCTTGAAGGTTCAGAGAACTTGGCTTCTGTGCGGTGCGGGGTGTGCGAG 960
QY 961 ATGGTACGAGACGTGCTGCGGTGACTCATCTGCGCGCTGCTTCCACTGAGCTGCAC 1020
DB 961 ATGGTACGAGACGTGCTGCGGTGACTCATCTGCGCGCTGCTTCCACTGAGCTGCAC 1020
QY 1021 TCCCAAGCCGACCTCCCGGCGCGAGCGAGGCTGGCTGACAGATCTGCTCAGAGAG 1080
DB 1021 TCCCAAGCCGACCTCCCGGCGCGAGCGAGGCTGGCTGACAGATCTGCTCAGAGAG 1080
QY 1081 TGAACCCAGCCCTGTGGAGGGGGGTGTCGCCCCAGCCCCCGGCTGGCGCCCTGG 1140
DB 1081 TGAACCCAGCCCTGTGGAGGGGGGTGTCGCCCCAGCCCCCGGCTGGCGCCCTGG 1140
QY 1141 CTGCCAAGATGACACTGCACTGCACTGCAAGACCCGCTCTGCAAGAGATGACTGAG 1200
DB 1141 CTGCCAAGATGACACTGCACTGCACTGCAAGACCCGCTCTGCAAGAGATGACTGAG 1200
QY 1201 TTCTGAGCAGACACCTTGTGATGTCATCTGCAAGTGGGCTGCAAGATGAGCCG 1260
DB 1201 TTCTGAGCAGACACCTTGTGATGTCATCTGCAAGTGGGCTGCAAGATGAGCCG 1260
QY 1261 CGCGGCGCCCTTCCCTCCTGACCCCAATGGCGGAGCATGAGCTGTGAGAGAG 1320
DB 1261 CGCGGCGCCCTTCCCTCCTGACCCCAATGGCGGAGCATGAGCTGTGAGAGAG 1320
QY 1321 TGTGAGAGAGACACTCTCTCTCTCACTCTGGAAGCCGAGCGGCTGGATCAAGAG 1380
DB 1321 TGTGAGAGAGACACTCTCTCTCTCACTCTGGAAGCCGAGCGGCTGGATCAAGAG 1380
QY 1381 GACAGCGCACTCTTGTGATGCTGCTGCTGTAACAGCTGTGTGTTCTGGGACAC 1440
DB 1381 GACAGCGCACTCTTGTGATGCTGCTGCTGTAACAGCTGTGTGTTCTGGGACAC 1440
QY 1441 GGCATGATGCTGGAATTAACCCCTGCCCACTTCTCTCTGGAAGTCCCGGGA 1500
DB 1441 GGCATGATGCTGGAATTAACCCCTGCCCACTTCTCTCTGGAAGTCCCGGGA 1500
QY 1501 GCTCTCTTGGCTGTGACTTAATAAATTAATAAATTAAGCT 1545
DB 1501 GCTCTCTTGGCTGTGACTTAATAAATTAATAAATTAAGCT 1545
QY 1419 GCTCTCTTGGCTGTGACTTAATAAATTAATAAATTAAGCT 1463
DB 1419 GCTCTCTTGGCTGTGACTTAATAAATTAATAAATTAAGCT 1463

RESULT 3
AAX26936
ID AAX26936 standard; DNA; 2036 BP.
XX
AC AAX26936;
XX
DT 24-JUN-1999 (first entry)

XX CDNA encoding a human autoimmune regulator-1 (AIR-1) protein.
DE
XX Autoimmune regulator-1; AIR-1; immune maturation; immune response;
XX disease; autoimmune polyendocrinopathy candidiasis ectodermal dys trophy;
KW APECED; autoimmune polyglandular syndrome type I; APS I; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 137..1774
FT /*tag= a "AIR-1"
FT /product= "AIR-1"
XX
XX WO915559-A1.
XX
XX 01-APR-1999.
XX
XX 23-SEP-1998; 98WO-FI000749.
XX
XX 23-SEP-1997; 97FI-00003762.
XX
XX (FIM-) FINNISH IMMUNOTECHNOLOGY LTD.
XX
XX Krohn K, Heino M, Peterson P, Scott H, Antonarakis S, Lalioti M,
PI Shintzu N, Kudoh J;
XX
XX WPI; 1999-244390/20.
XX
XX P-PSDB; AAY01712.
XX
XX Autoimmune regulator 1 (AIR1) DNA sequence.
XX
XX Claim 4; Page 21-24; 59pp; English.
XX
XX The present sequence encodes an autoimmune regulator-1 (AIR-1) protein.
XX The AIR polypeptides and polynucleotides can be used in methods for the
XX diagnosis and treatment of diseases related to immune maturation and
XX regulation of immune response towards self and nonself. They can be used
XX particularly in the diagnosis and treatment of autoimmune
XX polyendocrinopathy candidiasis ectodermal dys trophy (APECED) (also known
XX as autoimmune polyglandular syndrome type I (APS I))
XX
XX Sequence 2036 BP; 359 A; 718 C; 649 G; 310 T; 0 U; 0 Other;
SQ
Query Match 66.3%; Score 1023.6; DB 2; Length 2036;
Best Local Similarity 99.6%; Pred. No. 1.1e-208;
Matches 1026; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 516 CTGTGCGCAAGAATATGAGAGCGAGTGTGCTGTGCGGAGCGGGAGCTCATCTGC 575
DB 1007 CTCACACAGAAAGATAGAGAGAGTGTGCTGTGCGGAGCGGGAGCTCATCTGC 1066
QY 576 TGTGACGGCTGCGCTCGGGCTTCCACCTGAGCTGCTGCTCGCTCGGGAGATC 635
DB 1067 TGTGACGGCTGCGCTCGGGCTTCCACCTGAGCTGCTGCTCGCTCGGGAGATC 1126
QY 636 CCCAGTGGAGCTGAGGTGCTTCAGCTGCTGAGGCAACAGTCCAGAGGTGACGCC 695
DB 1127 CCCAGTGGAGCTGAGGTGCTTCAGCTGCTGAGGCAACAGTCCAGAGGTGACGCC 1186
QY 696 CCGGAGAGAGAGCGCGCGCCGAGGCAACCGGTGAGACCGCGCTCCCGGGGGCTT 755
DB 1187 CCGGAGAGAGAGCGCGCGCCGAGGCAACCGGTGAGACCGCGCTCCCGGGGGCTT 1246
QY 756 AGGTGCGCGGAGAGAGGTGAGAGGTTCACCTGAGGAAACCTTACCGGCATGACAG 815
DB 1247 AGGTGCGCGGAGAGAGGTGAGAGGTTCACCTGAGGAAACCTTACCGGCATGACAG 1306
QY 816 ACTCTTGTCTACAGACCTGCGGCTCGGCTTCTGACAGCCCGCTGCCAGGTCTGAC 875
DB 1307 ACTCTTGTCTACAGACCTGCGGCTCGGCTTCTGACAGCCCGCTGCCAGGTCTGAC 1366
QY 876 TCTGTGCGCTGACACCCCTACTGTGTGTGGGTCTGAGAGGTGACAGAAACCTGCTCCT 935

DB 1367 TCTGTGCGCTGACACCCCTACTGTGTGTGGGTCTGAGAGGTGACAGAACTGCTCCT 1426
QY 936 GGTGCGCTTGTGCGGGGTGTGCGAGATGTAACGACGCTGCTGCTGTACTGACCTGCC 995
DB 1427 GGTGCGCTTGTGCGGGGTGTGCGAGATGTAACGACGCTGCTGCTGTACTGACCTGCC 1486
QY 996 GCTGCTTCTACTGCGCTGCGCACTTCCAGCGGCACTCCCGCGCGGAGACGGGCTTG 1055
DB 1487 GCTGCTTCTACTGCGCTGCGCACTTCCAGCGGCACTCCCGCGCGGAGACGGGCTTG 1546
QY 1056 CGCTGAGATCTGCTCAGAGAACGTAACCGAGCCCGCTGAGAGGGGGTGTGAGCCGCC 1115
DB 1547 CGCTGAGATCTGCTCAGAGAACGTAACCGAGCCCGCTGAGAGGGGGTGTGAGCCGCC 1606
QY 1116 AGCCCGCGCGCTGCGCTGCGCAAGATGACCTGCGCAAGTACAGACGCGCTG 1175
DB 1607 AGCCCGCGCGCTGCGCTGCGCAAGATGACCTGCGCAAGTACAGACGCGCTG 1666
QY 1176 CTGACAGGAGATGACCTGAGATCTCTTCTGACGAGACACCTTGAATGGCATCTGACG 1235
DB 1667 CTGACAGGAGATGACCTGAGATCTCTTCTGACGAGACACCTTGAATGGCATCTGACG 1726
QY 1236 TGGGCGATCAGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1295
DB 1727 TGGGCGATCAGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1786
QY 1296 GGGACATGAGCTGATGAGAGAGTGTGAGAGAGACACTTCTCTCACTGAGTAA 1355
DB 1787 GGGACATGAGCTGATGAGAGAGTGTGAGAGAGACACTTCTCTCACTGAGTAA 1846
QY 1356 GCGGCGCGCGCTGAGATCAAGAGGAGACAGCGCACCTTCTCACTGAGTAA 1415
DB 1847 GCGGCGCGCGCTGAGATCAAGAGGAGACAGCGCACCTTCTCACTGAGTAA 1906
QY 1416 CAGCTGTGTTCTGCGGAGACAGCGCATGAGCTGAGAAATTAACCTGCGCCAC 1475
DB 1907 CAGCTGTGTTCTGCGGAGACAGCGCATGAGCTGAGAAATTAACCTGCGCCAC 1966
QY 1476 TTCTTACTTGAAGTCCCGGAGACCTCTCTGCTGAGTAAATTAATTA 1535
DB 1967 TTCTTACTTGAAGTCCCGGAGACCTCTCTGCTGAGTAAATTAATTA 2026
QY 1536 AAATTAGCTG 1545
DB 2027 AAATTAGCTG 2036
RESULT 4
AAK58605
ID AAK58605 standard; cDNA; 2245 BP.
XX
XX AAK58605;
AC 16-AUG-1999 (first entry)
XX
XX Human autoimmune polyglandular disease type 1 (APGD1) cDNA.
DE
XX Autoimmune polyendocrinopathy candidiasis ectodermal dys trophy; APECED;
KW autoimmune polyglandular disease type 1; APGD1; AIRB; human;
KW transcription factor; autoimmune disease; diagnosis; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 121..1758
FT /*tag= a "human_APGD1"
FT /product= "human_APGD1"
FT replace(358,T)
FT /*tag= k
FT mutation
FT replace(374,G)
FT /*tag= 1
FT variation
FT replace(708,T)

FT		/tag= o	/note= "silent polymorphism"
FT		/replace(801,T)	
FT	variation	/tag= p	/note= "silent polymorphism"
FT		/replace(889,T)	
FT	mutation	/tag= j	/note= "alters Arg codon to STOP, and truncated 256 aa protein"
FT		1051	
FT	mutation	/tag= g	/note= "1-nucleotide deletion"
FT		/replace(1052,A)	
FT	mutation	/tag= m	
FT		1085..1097	
FT		/tag= f	/note= "13-nucleotide deletion, results in frameshift and truncated 372 aa protein"
FT	mutation	/replace(1085,CCTG)	
FT		/tag= b	/note= "4-nucleotide insertion, results in frameshift and truncated 371 aa protein"
FT	mutation	/replace(1090,CCTG)	
FT		/tag= c	/note= "4-nucleotide insertion, results in frameshift and truncated protein"
FT	mutation	/replace(1094,A)	
FT		/tag= n	/replace(1284,A)
FT	mutation	/tag= d	/note= "1-nucleotide insertion, results in frameshift and truncated 422 aa protein"
FT	mutation	/tag= h	/note= "1-nucleotide deletion"
FT		1313	
FT	mutation	/tag= i	/note= "1-nucleotide deletion, results in frameshift and truncated 478 aa protein"
FT	variation	/replace(1317,T)	
FT		/tag= q	/note= "silent polymorphism"
FT	mutation	/replace(1365,C)	
FT		/tag= e	/note= "1-nucleotide insertion"
FT	variation	/replace(1698,T)	
FT		/tag= r	/note= "silent polymorphism"
FT	misc_feature	1995	
FT		/tag= s	/note= "Alu sequence overlapping with PFKL promoter starts at this position"
FT	polyA_signal	2191..2195	
FT		+tag= t	
XX			
PN	W09918197-A2.		
XX			
XX	15-APR-1999.		
PD			
XX			
PP	02-OCT-1998;	98WO-EP006294.	
XX			
PR	02-OCT-1997;	97EP-00117154.	
PR	08-OCT-1997;	97EP-00117398.	
PR	12-NOV-1997;	97EP-00119810.	
XX			
PA	(PLAC) MAX PLANCK GRS FOERDERUNG WISSENSCHAFTEN.		
PA	(NAPU-) NAT PUBLIC HEALTH INST.		
XX			
PI	Peltonen L, Aaltonen J, Bjorsetes P, Perheentupa J, Palotie A;		
PI	Horelli-Kuitunen N, Yaaipo M, Lehnach H;		
XX			
XX	WP1; 1999-287735/24.		
RR	P-PSDB; AAAY06073.		

XX
PT New polypeptide which co-segregates in mutated form.
PS
PP Claim 1; Fig 2A; 77pp; English.

CC This is the nucleotide sequence of human autoimmune polyglandular disease type 1 (APGDI or AIRE) cDNA. It is a composite of 2 clones, which were obtained from a human adult thymus cDNA library by PCR (see AA58609-10), and of a 3'UTR extension PCR product. The encoded protein (see AA586073) is a transcription factor or transcription-associated factor that may associate with vimentin fibres, perhaps as part of a docking mechanism regulating nuclear translocation. Aggregates of the mutated protein may prevent formation of vimentin intermediate filaments. Mutated APGDI polypeptides co-segregate with autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED). The invention provides vectors and host cells for preparation of APGDI polypeptides. APECED-associated mutations found in the APGDI gene include: an insertion of 4 nucleotides (CCTG)_n, normally found at position 1086-1089, at nucleotide position 1085 or 1090 ; an insertion of an adenosine at position 1284; an insertion of a cytosine at position 1365; a deletion of nucleotides 1085-1097; a deletion of the thymidine at position 1051; a deletion of the cytosine at position 1309 or 1313; a cytosine to thymidine exchange at position 889; a guanosine to thymidine exchange at position 358; an adenosine to a guanosine exchange at position 374; a guanosine to adenosine exchange at position 1052; and a cytosine to adenosine exchange to position 1094. A claimed method for testing for carriage of APECED or for the disease state involves testing for a mutation in the APGDI gene or for a mutated form of the APGDI polypeptide. APGDI nucleic acid can be used in gene therapy

Query Match 66.3%; Score 1023.6; DB 2; Length 2245;

```
Matches 1026; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

QY	516	CTGTGTCACAGAGAA	TAGAGACGAGTGTGTC	TGTGTGCGGACCGCGGGAGCTCAT	TCG	575
Db	991	CTCCACCCAGGAAGAT	TGAGACGAGTGTGTC	GTGTGTCGGACCGCGGGAGCTCAT	TCG	1050

Qy 576 TGTACGCGCTGCCTTCGGGCTTTCACCTGAGCTGACCTTGCCCTTCGGCTCCGGAGATC 635

Dp 1051 TGTACGCGCTGCCTTCGGGCTTTCACCTGAGCTGACCTTGCCCTTCGGCTCCGGAGATC 1111

636 CCGAGTGGGACCTGGAGTGTCTCCAGCTGCTCGAGGCAACAGTCCAGGAGGTGCAGCCC 695

696 CGGGCAGAGGAGCCCGGCCCCCAGAGCCACCGCTGAGAGCCCGCCCTCCCCCGGGCTTT 755

756 AGTGGCGGGGAGGAGGTAAAGGTCCACTGGGGAAACCCCTAAGCCGGCATGACAG 815

DB 123 ATGTCGCGCGAGAGAGGATAGAGAGTCTCAGCTTGGGAGATCCCTTAGCGCGGCTGGACATCG 123

QY 816 ACTCTGTCTACAAGACACCTGCCGGCTTCGCGCTTCTGCAAGCCCGCGCTGCAGGTCTGGAC 875

Db	1291	ACTCTGTCTACAGACACCTGCCGGCTCCGCTTCTGCAGGCCCGCTGCCAGGGCTGCAC	1356
Qy	876	TCTCTGGCCCTGCACCCCTACTGTGTGTGGTCTCTGAAGGTCAGACAGACTGGCTCT	935

Db 1351 TCCTCGGCGCTGCAACCCCTACTGTGTGTGGGTCTCTGAGGGTCAGACAGAACCTGGCTCCT 1410

Qy 936 GGTGCGCGCTTTGCGGGGTGTGCGGAGATGTACGGAACGTCTGCCGTGTACTCACTGGCGC 995

Db 1411 GGTGGCGCTGTGCGGAGATGTAACGACGTCTGCGGTGTACTACATGCGCC 1476

Db	1471	GCTGCCTTCACGTGGCGCTGCCATTCCAGCGCCGCACTCCCGCCCGGGAGCGGGCTTG	153
QY	1056	CGCTGCAGATCTCTGCTCAGAGAGCGTGACCCAGCCGCCCTGTGTGAGGGGGGTGTGGCCCC	111

DB 1587 TCACGGCATCTCGACTGGGCGCATCCAGCATGTCCAGCCCGCTGCGGAGACACACC 1646
 QY 1280 CT 1281
 DB 1647 CT 1648

RESULT 6
 ABL32092
 ID ABL32092 standard; DNA; 9876 BP.

AC ABL32092;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 65.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-anaemic; cyrostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

PS Claim 1, SEQ ID NO 65; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX Sequence 9876 BP; 1912 A; 309 C; 2928 G; 4727 T; 0 U; 0 Other;

Query Match 7.7%; Score 118.4; DB 6; Length 9876;
 Best Local Similarity 76.0%; Pred. No. 1.6e-15;
 Matches 146; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1 AGAGAAAGTGAAGCTCTTCTCAGGCTTTAAGACATGGCGTTTGTCCAGGCTGTAACCG 60
 DB 9386 AAGAGAAAGTGAAGCTCTTCTCAGGCTTTAAGAGATATGGGTTTGTAGTTATTCG 9445
 QY 61 CTGCTCTCAGCTGGGCGCGTGGTGGCGGGCGCCCTGCTATACCGAGAGGTCAAG 120
 DB 9446 TTGTTTATAGTTGGGTTTGTGGTGGTGGGCGGCTTTTGTATTAAGTTAGAGAGTTAAG 9505

QY 121 ATCCATGGGAATGCATGCTCATCTTTCGTCGCCAGCATGTTCTTAATGGGTAAGA 180
 DB 9506 ATTATTTGGAAATGTTATGTTATTTTCGTTTAAATGATGTTTAAATGGGTAAGA 9565
 QY 181 GCAGGTGGGAG 192
 DB 9566 GTAGTGTGGCGG 9577

RESULT 7
 ABL32093/C
 ID ABL32093 standard; DNA; 9876 BP.

AC ABL32093;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 66.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-anaemic; cyrostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

PS Claim 1, SEQ ID NO 66; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX Sequence 9876 BP; 1882 A; 309 C; 3154 G; 4531 T; 0 U; 0 Other;

Query Match 6.6%; Score 101.8; DB 6; Length 9876;
 Best Local Similarity 69.2%; Pred. No. 5.6e-12;
 Matches 139; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 AGAGAAAGTGAAGCTCTTCTCAGGCTTTAAGACATGGCGTTTGTCCAGGCTGTAACCG 60
 DB 491 AAAAAAATAAATTTCTCAAACTTAATAAACAATAGTTTAATCCAACTATACCG 432
 QY 61 CTGCTCTCAGCTGGGCGCGTGGTGGCGGGCGCCCTGCTATACCGAGAGGTCAAG 120
 DB 431 CTACTCTCAACTAAACCGTAATAAATTAACCGAACGCCCTTAATTAACCAAAAAATCAAAA 372

QY 121 ATCCACTGGGAATGCCATCTCATCTTTGTCCTCCGACATGGTTCTTAATGGGTAGAA 180
 DB 371 ATCCACTGAATAAATACCATCTACTCTTGTGCTCCCAACATATTTCTTAATAAATAA 312
 QY 181 GCAGGTGCGGAGAGACCTGCC 201
 DB 311 ACATATATATAAATACCTACC 291

RESULT 8
 ADH12179
 ID ADH12179 standard; cDNA; 9646 BP.
 AC ADH12179;
 DT 11-MAR-2004 (first entry)
 DE Human CHD5 (chromodomain, helicase, DNA-binding 5)-encoding cDNA.
 XX
 KW Human; chromodomain helicase DNA-binding 5; CHD5; chromosome 1p36.3;
 KW chromatin structure; chromatin unwinding; DNA repair; DNA recombination;
 KW transcriptional regulation; gene expression; cell cycle control;
 KW development regulation; oncogenesis; brain; neural development;
 KW neural tissue neoplasia; diagnosis; cancer; neural cancer; neuroblastoma;
 KW breast cancer; colon cancer; liver tumour; germ cell tumour;
 KW drug screening; cytostatic; gene therapy; gene; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 101..5965
 FT /tag=a
 FT /product="CHD5"
 XX
 PN WO2003106650-A2.
 PD 24-DEC-2003.
 XX
 PF 16-JUN-2003; 2003WO-US019027.
 XX
 PR 14-JUN-2002; 2002US-0388848P.
 PA (CHIL-) CHILDERNS HOSPITAL PHILADELPHIA.
 PI Brodeur GW, White PS;
 DR MPI; 2004-082186/08.
 DR P-PsDB; ADH12180.
 PT Novel chromodomain helicase DNA-binding (CHD) proteins, useful as
 PT diagnostic and prognostic indicator of tumor, comprises amino terminus
 PT having two PHD class zinc finger domains and two chromodomains.
 XX
 PS Claim 5; SEQ ID NO 1; 124pp; English.

XX The invention relates to human chromodomain, helicase, DNA-binding 5
 CC (CHD5; ADH12180) and cDNA encoding it (ADH12179). CHD5 is a novel member
 CC of the CHD gene family, members of which have a profound effect on
 CC chromatin structure and gene expression and which are thus likely to play
 CC an important role in cell cycle control, regulation of development, and
 CC oncogenesis. CHD5 comprises two N-terminal zinc finger domains of the PHD
 CC (plant homeodomain) class, two chromodomains, a central region which
 CC contains a predicted DEAH-box-type helicase domain and a putative SNF2
 CC domain, and several nuclear localization signals. The gene encoding CHD5
 CC is located on chromosome 1p36.3, a region frequently deleted in a variety
 CC of cancers including neuroblastoma, and the protein is preferentially
 CC expressed in brain. CHD5 is therefore thought to be a modulator of normal
 CC neural development and neoplasias of neural tissue origin. The invention
 CC also relates to vectors and host cells comprising the CHD5 cDNA sequence;
 CC an antibody against CHD5; a method of screening for modulators of CHD5
 CC activity; a method of diagnosing cancer in a patient, where a reduced
 CC level or absence of CHD5 or CHD5 nucleic acids indicates the presence of

CC cancer; treating cancer by administration of CHD5 protein, CHD5-encoding
 CC nucleic acids or CHD5 mimetics; and CHD5-specific PCR primers (ADH12186-
 CC ADH12247). The methods of the invention are useful in the diagnosis or
 CC treatment of cancers such as neural cancers (e.g., neuroblastoma), breast
 CC cancer, colon cancer, liver tumours and germ cell tumours. The CHD5
 CC protein, CHD5 nucleic acids and anti-CHD5 antibodies are useful as
 CC research tools to identify other proteins that are intimately involved in
 CC chromatin unwinding, DNA repair and recombination, and transcriptional
 CC regulation. The present sequence represents cDNA encoding human CHD5.
 XX
 SQ Sequence 9646 BP; 2076 A; 2789 C; 3013 G; 1768 T; 0 U; 0 Other;
 Query Match 4.2%; Score 65; DB 12; Length 9646;
 Best Local Similarity 62.7%; Pred. No. 0.0004;
 Matches 101; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTCCCGTGTGCGGACGGGGGAGAGTCTGTGTCAGGCTGCCCTC 591
 DB 1128 AGGATTACTGTAGTGTGCTCCAGCAGGTGGGAGATCTCTGTCCGACACTGCCCA 1187
 QY 592 GGGCTTCACCTGTGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGAGACTGGA 651
 DB 1188 GGGCTTACCATCTCGTATGCTGTCGACCCAGACTGGAGAAAGGCTCCGAGGCAAGTGA 1247
 QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGGTGAG 692
 DB 1248 GCTGCCCCCACTGTGAGAGAGGGGATCAGTGGAGCCG 1288

RESULT 9
 ABN38615
 ID ABN38615 standard; DNA; 60 BP.
 AC ABN38615;
 XX
 DT 15-JUL-2002 (first entry)
 DE Human spliced transcript detection oligonucleotide SEQ ID NO:11363.
 XX
 KW Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcriptome; oligonucleotide library; ss.
 OS Homo sapiens.
 PN WO200210449-A2.
 PD 07-FEB-2002.
 XX
 PF 20-JUL-2001; 2001WO-1B001903.
 PR 28-JUL-2000; 2000US-0221607P.
 PR 02-MAY-2001; 2001US-0287724P.
 PA (COMP-) COMPUGEN INC.
 PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 DR MPI; 2002-257383/30.
 PT New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of a
 PT genome, useful for detecting tissue-, pathology-, and developmental-
 PT specific genes.
 XX
 PS Example 1; SEQ ID NO 11363; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-
 CC)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises several
 CC oligonucleotides, each capable of hybridizing selectively to a set of
 CC messenger RNAs transcribed from a given transcription unit of the genome,
 CC which encodes one or more messenger RNA splice variants. The

XX (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
 XX Seelig HP, Renz M;
 XX WPI: 1996-240280/25.
 DR P-PSDB; AAR99334.
 XX
 PT DNA encoding dermatomyositis specific auto:antigen - useful for
 XX differential diagnosis and treatment of dermatomyositis.
 PS Claim 3; Fig 2; 20pp; German.
 XX
 CC The present sequence encodes a 218 kD dermatomyositis specific auto-
 CC antigen, designated M1-2. The sequence numbering given in the
 CC specification starts at nucleotide 91, i.e. the 5' UTR is omitted, as
 CC well as the 'A' of the first ATG start codon. The protein is hydrophilic,
 CC acidic and protruding regions of the protein are characteristic of
 CC helices. The gene corresponding to this cDNA was localised to
 CC chromosome 12 (12p13). The DNA can be used for the recombinant production
 CC of M1-2 which is used for, e.g. the differential diagnosis of collagen
 CC diseases, esp. dermatomyositis, e.g. by immunoassay or Western blotting
 XX
 SQ Sequence 6327 BP; 1761 A; 1473 C; 1740 G; 1353 T; 0 U; 0 Other;
 Query Match 3.9%; Score 60; DB 2; Length 6327;
 Best Local Similarity 61.5%; Pred. No. 0.0044;
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 532 AGGACGAGTGTCCCGTGTCTCGGAGCGGGAGCTCATCTGCTGACGCTGCCCTC 591
 DB 1108 AGGACTATTGCGAGGTGTCGACGAGCGGTGATCATCTGTGATCACTGTCC 1167
 QY 592 GGGCCCTTCACGCTGCGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGAGCTGGA 651
 DB 1168 GTGCTTACCACTGATGCTGCTGCTGATCCGACATGAGAGAGCTCCCGAGGGAATGGA 1227
 QY 652 GGTGCTCCAGCTGCTGTCAGGCAACAGTCCAGAGG 687
 DB 1228 GCTGCCACACTGCGAGAGGAGCATCCAGTGGG 1263
 RESULT 12
 ABL64410
 ID ABL64410 standard; DNA; 6417 BP.
 XX
 AC ABL64410;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Stomach cancer related gene sequence SEQ ID NO:2747.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytotoxic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200194629-A2.
 PD 13-DEC-2001.
 PF 30-MAY-2001; 2001WO-US010838.
 XX
 PR 05-JUN-2000; 2000US-0209473P.
 PR 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0231337P.
 PR 18-SEP-2000; 2000US-0233617P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234509P.

PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 PA (AVAL-) AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 DR WPI: 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 PS Claim 1; SEQ ID NO 2747; 44pp; English.
 XX
 CC The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (II) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' s
 CC tumour
 XX
 SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;
 Query Match 3.9%; Score 60; DB 6; Length 6417;
 Best Local Similarity 61.5%; Pred. No. 0.0044;
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 532 AGGACGAGTGTCCCGTGTCTCGGAGCGGGAGCTCATCTGCTGACGCTGCCCTC 591

DB 1198 AGAGCTATTGCCAGGCTGTGCAGCAAGGGGTGAGATCATCCGTGTGTAACCTGTCCCC 1257
QY 592 GGGGCTTCCACCTGCGCTGCTCTCCCTCCGTCGGGAGATCCCAAGTGGAGACTGTGA 651
DB 1258 GTGCTTACCAATGCTGTGCTGTGATCCCGACATGGAAGAGGCTCCCGAGGCAAGTGA 1317
QY 652 GGTGCTCCAGCTGCTGTCAGGCAAGCATGTCAGGAGG 687
DB 1318 GCTGCCACACTGCGAAGGAAGGCAATCCATGCGG 1353

RESULT 13
ABX04169
ID ABX04169 standard; cDNA; 6417 BP.
AC ABX04169;
XX
XX 10-JAN-2003 (first entry)
DE Human mRNA differentially expressed in mesenchymal cells #16.
XX
XX Human; ss; gene; skeletal growth; cartilage degeneration disorder;
KM chondroblastic phenotype; mesenchymal cell; cartilage formation;
KM bone formation; arthritis; osteoarthritis; rheumatoid arthritis;
KM gout arthritis; adjuvant arthritis; arthritis deformans; ankylosis;
KM infectious arthritis; osteochondrosis; RDA; antihypertic; osteopathic;
KM antineumatic; antiinflammatory; representational difference analysis.
XX
XX Homo sapiens.
OS
XX WO200271927-A2.
PN
XX 19-SEP-2002.
PD
XX 12-MAR-2002; 2002WO-US007787.
PF
XX 12-MAR-2001; 2001US-0274980P.
PR
XX (BGMH) BRIGHAM & WOMENS HOSPITAL INC.
PA
XX Yates KE, Mizuno S, Glowacki J;
PI
XX MPI; 2002-723276/78.
DR
XX
PT New nucleic acid molecules capable of promoting chondrogenesis, useful
PT for diagnosing and treating cartilaginous tissue degeneration conditions,
PT e.g. osteoarthritis, rheumatoid arthritis, gout arthritis, or
PT osteochondrosis.
PS
XX Claim 33; Page 111-113; 153pp; English.
XX

The invention relates to new isolated nucleic acid molecule comprising a nucleic acid molecule consisting of a gene differentially expressed in cells undergoing differentiation from mesenchymal cell to a chondroblastic phenotype, or hybridising under stringent conditions to them (or their fragments). Also included are expression vectors, transformed host cells, expressed polypeptides or peptide fragments (which induce differentiation of a mesenchymal cell and may be used as an immunogen), binding partners of the polypeptides, a method for identifying an agent useful in modulating mesenchymal cell differentiation induction activity of a molecule, a method of diagnosing a condition characterized by aberrant expression of a nucleic acid molecule or its expression product; a method for determining regression, progression or onset of cartilaginous tissue degeneration condition in a subject characterised by aberrant expression of a nucleic acid molecule or its expression product, a method for treating a cartilaginous tissue degeneration condition, a method for treating a subject to reduce the risk of cartilaginous tissue degeneration condition developing in the subject, a method for identifying a candidate agent for treating a cartilaginous tissue degeneration condition, and a solid-phase nucleic acid molecule array consisting essentially of a set of nucleic acid molecule as cited above (or known from known genes shown to be

CC differentially expressed in developing mesenchymal cells using the CC technique of representational difference analysis, RDA), its expression CC products or fragments, fixed to a solid substrate. The nucleic acids, CC polypeptides and agents are useful for treating cartilaginous tissue CC degeneration conditions such as osteoarthritis, rheumatoid arthritis, CC gout arthritis, adjuvant arthritis, arthritis deformans, infectious CC arthritis or osteochondrosis. The present sequence is a cDNA from a CC known gene differentially expressed in developing mesenchymal cells XX
SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;
Query Match 3.9%; Score 60; DB 6; Length 6417;
Best Local Similarity 61.5%; Pred. No. 0.0044;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGAGCAATGTCGCTGTCGCGACGCGGGGAGCTCATCTGCTGTGACGGCTGCCCTC 591
DB 1198 AGAGCTATTGCCAGGCTGTGCAGCAAGGGGTGAGATCATCCGTGTGTAACCTGTCCCC 1257
QY 592 GGGGCTTCCACCTGCGCTGCTCTCCCTCCGTCGGGAGATCCCAAGTGGAGACTGTGA 651
DB 1258 GTGCTTACCAATGCTGTGCTGTGATCCCGACATGGAAGAGGCTCCCGAGGCAAGTGA 1317
QY 652 GGTGCTCCAGCTGCTGTCAGGCAAGCATGTCAGGAGG 687
DB 1318 GCTGCCACACTGCGAAGGAAGGCAATCCATGCGG 1353

RESULT 14
ABK84696
ID ABK84696 standard; cDNA; 6417 BP.
AC ABK84696;
XX
XX 14-AUG-2002 (first entry)
DE Human cDNA differentially expressed in granulocytic cells #1267.
XX
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KM viral infection; parasitic infection; protozoal infection;
KM fungal infection; sterile inflammatory disease; psoriasis;
KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KM cardiac reperfusion injury; renal reperfusion injury; ARDS;
KM adult respiratory distress syndrome; inflammatory bowel disease;
KM Crohn's disease; ulcerative colitis; periodontal disease;
KM granulocyte activation; chronic inflammation; allergy.
XX
XX Homo sapiens.
OS
XX WO200228999-A2.
PN
XX 11-APR-2002.
PD
XX 03-OCT-2001; 2001WO-US030821.
PF
XX 03-OCT-2000; 2000US-0237189P.
PR
XX (GENE-) GENE LOGIC INC.
PA
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
PI
XX MPI; 2002-435328/46.
DR
XX
PT Detecting granulocyte activation by detecting differential expression of PT genes associated with granulocyte activation, which serves as diagnostic PT markers that is useful for monitoring disease states and drug toxicity.
PS
XX Claim 1; SEQ ID NO 1267; 114pp; English.
XX

The invention relates to detecting (M) granulocyte (GC) activation CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by CC DNA chip analysis as given in the specification, and comparing the CC expression level to an expression level in an unactivated GC, where

CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease, also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

CC Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;

CC Query Match 3.9%; Score 60; DB 6; Length 6417;

CC Best Local Similarity 61.5%; Pred. No. 0.0044; Mismatches 60; Indels 0; Gaps 0;

CC Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGTGACGCGTGCCTTC 591
DB 1198 AGGACTATTGCGAGGTGTGCGGACGCGGTGATCATCTGTGTATACCTGTCCTCC 1257
QY 592 GGGCCCTTCACCTGCGCTGCTCTCCCTCCGCTCCGCGAGATCCCAAGTGGAGCTTGA 651
DB 1258 GTGCTTACCACTGCTGTGCTGTGATCCCGACATGGAAGAGCTCCCGAGGCAAGTGA 1317
QY 652 GGTGCTCCAGCTGCTGCGAGCAACAGTCCAGAGG 687
DB 1318 GCTGCCACACTGCGAGGAAGGATCCAGTGGG 1353

CC RESULT 15

CC AADS4631

CC ID AADS4631 standard; DNA; 6417 BP.

CC AC AADS4631;

CC DT 26-JUN-2003 (first entry)

CC DE Human chromodomain helicase DNA binding protein (CHD) encoding DNA #6.

CC KM Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;

CC KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;

CC KM cell proliferative disorder; chromatin organisation modifier domain;

CC XX cytoskeletal; de.

CC OS Homo sapiens.

CC XX MO200298899-A2.

CC PD 12-DEC-2002.

CC XX 03-JUN-2002; 2002MO-US017466.

XX 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 22-OCT-2001; 2001US-0338733P.
PR 15-FEB-2002; 2002US-0357253P.
PR 15-FEB-2002; 2002US-0357600P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

PI Lionbin MN;

XX WPI; 2003-156840/15.

PT Identifying a candidate p53 pathway-modulating agent as therapeutic
PT targets for disorders related to defective p53 function e.g. cancer by
PT contacting an assay system having purified CHD polypeptide or nucleic
PT acid, with a test agent.

XX Disclosure; Page 59-62; 278pp; English.

XX The present invention relates to a method for identifying candidate p53
XX pathway modulating agents. The method involves contacting an assay system
XX comprising purified chromatin organisation modifier (chromo) domain
XX helicase DNA binding proteins (CHD), nucleic acids, their functionally
XX active fragments or derivatives, with a test agent under conditions
XX where, but for the presence of the test agent, the system provides a
XX reference activity. The methods are useful for identifying modulators of
XX the p53 pathway as therapeutic targets for disorders associated with
XX defective p53 function, such as angiogenic disorders, apoptotic disorders
XX or cell proliferative disorders, e.g. cancer. The modulators are useful
XX as research reagents, diagnostics and therapeutics. The invention is also
XX useful in gene therapy. The present sequence is human CHD DNA

XX Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;

XX Query Match 3.9%; Score 60; DB 10; Length 6417;

XX Best Local Similarity 61.5%; Pred. No. 0.0044; Mismatches 60; Indels 0; Gaps 0;

XX Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGTGACGCGTGCCTTC 591
DB 1198 AGGACTATTGCGAGGTGTGCGGACGCGGTGATCATCTGTGTATACCTGTCCTCC 1257
QY 592 GGGCCCTTCACCTGCGCTGCTCTCCCTCCGCTCCGCGAGATCCCAAGTGGAGCTTGA 651
DB 1258 GTGCTTACCACTGCTGTGCTGTGATCCCGACATGGAAGAGCTCCCGAGGCAAGTGA 1317
QY 652 GGTGCTCCAGCTGCTGCGAGCAACAGTCCAGAGG 687
DB 1318 GCTGCCACACTGCGAGGAAGGATCCAGTGGG 1353

CC RESULT 16

CC AADS4632

CC ID AADS4632 standard; DNA; 6417 BP.

CC AC AADS4632;

CC DT 26-JUN-2003 (first entry)

CC DE Human chromodomain helicase DNA binding protein (CHD) encoding DNA #7.

CC KM Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;

CC KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;

CC KM cell proliferative disorder; chromatin organisation modifier domain;

CC XX cytoskeletal; de.

CC OS Homo sapiens.

CC XX MO200298899-A2.

CC PD 12-DEC-2002.

CC XX 03-JUN-2002; 2002MO-US017466.

XX 03-JUN-2002; 2002MO-US017466.
PF
XX
PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 22-OCT-2001; 2001US-038733P.
PR 15-FEB-2002; 2002US-0357253P.
PR 15-FEB-2002; 2002US-0357600P.
XX
XX (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP,
PI Lioubin MN;
XX
XX MPI; 2003-156840/15.
XX
XX Identifying a candidate p53 pathway-modulating agent as therapeutic
PT targets for disorders related to defective p53 function e.g. cancer by
PT contacting an assay system having purified CHD polypeptide or nucleic
PT acid, with a test agent.
XX
XX
XX Disclosure; Page 62-66; 278bp; English.
XX
XX The present invention relates to a method for identifying candidate p53
CC pathway modulating agents. The method involves contacting an assay system
CC comprising purified chromatin organisation modifier (chromo) domain
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally
CC active fragments or derivatives, with a test agent under conditions
CC where, but for the presence of the test agent, the system provides a
CC reference activity. The methods are useful for identifying modulators of
CC the p53 pathway as therapeutic targets for disorders associated with
CC defective p53 function, such as angiogenic disorders, apoptotic disorders
CC or cell proliferative disorders, e.g. cancer. The modulators are useful
CC as research reagents, diagnostics and therapeutics. The invention is also
CC useful in gene therapy. The present sequence is human CHD DNA
XX
SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;
XX
Query Match 3.9%; Score 60; DB 10; Length 6417;
Best Local Similarity 61.5%; Pred. No. 0.0044;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTGCGGTGTGTGCGGACGCGGGAGCTCATCTGCTGTGACGGCTGCGCTC 591
DB 1198 AGGACTATTGCGAGGTGTGCGGACGAGGCGGTGATCATCTGTGTATCACTGTCC 1257
QY 592 GGGCCTTCCACCTGCGCTGCTGTCCCTCCCTCCGCGGAGATCCCAAGTGGACCTGGA 651
DB 1258 GTGCTTACCACTGATGTGCTGTGATCCGACATGGAAGGCTCCCGAGGCGCAAGTGA 1317
QY 652 GGTGCTCCAGCTGCTGCGGACGACAGATCCAGAGG 687
DB 1318 GCTGCCACACTGCGAGAAAGGATCCAGTGGG 1353

RESULT 17
ADP10531
ID ADP10531 standard; DNA; 6417 BP.
XX
XX ADP10531;
XX
XX 12-AUG-2004 (first entry)
XX
XX
DE Reference mRNA sequences for marker probe #208.
XX
XX transplant rejection; immune system; rheumatoid arthritis; lupus;
KW inflammatory bowel disease; multiple sclerosis; HIV; AIDS; ss.
XX
XX Homo sapiens.
XX
XX WO2004042346-A2.
XX
XX 21-MAY-2004.
XX

XX 24-APR-2003; 2003MO-US012946.
PF
XX
XX 24-APR-2002; 2002US-00131831.
PR 20-DEC-2002; 2002US-00325899.
XX
XX (EXPR-) EXPRESSION DIAGNOSTICS INC.
XX
XX Mohlgemuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M,
PI Rosenberg S;
XX
XX MPI; 2004-400724/37.
XX
XX
XX Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,
PT pancreas, pancreatic islet, lung, bone marrow or stem cell transplant
PT rejection, in an individual, comprises detecting the expression level of
PT the genes.
XX
XX
XX Claim 80; SEQ ID NO 540; 1762bp; English.
XX
XX
XX The present invention relates to diagnosing or monitoring transplant
CC rejection, e.g. cardiac or kidney transplant rejection, in an individual
CC comprising detecting the expression level of one or more genes. The
CC methods, system and kits are useful in diagnosing or monitoring
CC transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic
CC islet, lung, bone marrow or stem cell transplant rejection,
CC xenotransplant rejection or mechanical organ replacement rejection, in an
CC individual. The method is also useful in assessing the immune status of
CC an individual. The methods are also useful in diagnosing and monitoring
CC diseases that involve the immune system, e.g. rheumatoid arthritis,
CC lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or
CC viral, bacterial or fungal infection. The present sequence represents a
CC mRNA reference sequence for a 50 mer oligonucleotide marker for diagnosis
CC and monitoring of allograft rejection and other disorders.
XX
SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;
XX
Query Match 3.9%; Score 60; DB 12; Length 6417;
Best Local Similarity 61.5%; Pred. No. 0.0044;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTGCGGTGTGTGCGGACGCGGGAGCTCATCTGCTGTGACGGCTGCGCTC 591
DB 1198 AGGACTATTGCGAGGTGTGCGGACGAGGCGGTGATCATCTGTGTATCACTGTCC 1257
QY 592 GGGCCTTCCACCTGCGCTGCTGTCCCTCCCTCCGCGGAGATCCCAAGTGGACCTGGA 651
DB 1258 GTGCTTACCACTGATGTGCTGTGATCCGACATGGAAGGCTCCCGAGGCGCAAGTGA 1317
QY 652 GGTGCTCCAGCTGCTGCGGACGACAGATCCAGAGG 687
DB 1318 GCTGCCACACTGCGAGAAAGGATCCAGTGGG 1353

RESULT 18
AA158444
ID AA158444 standard; cDNA; 6475 BP.
XX
XX AA158444;
XX
XX 22-OCT-2001 (first entry)
XX
XX
DE Human polynucleotide SEQ ID NO 647.
XX
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX
XX Homo sapiens.
XX
XX

XX ss; cancer; neurodegenerative disease; human.
KM Homo sapiens.
OS US2003104529-A1.
PN XX
PD 05-JUN-2003.
XX
PF 04-JAN-2002; 2002US-00037270.
XX
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 19-JUL-2000; 2000US-00620312.
XX
PA (ZHOU/) ZHOU P.
PA (TANG/) TANG Y T.
PA (LIU/) LIU C.
PA (ASUNDI/) ASUNDI V.
PA (DRMANAC/) DRMANAC R T.
XX
PI Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;
XX
DR WPI; 2003-678194/64.
XX
PT New polynucleotide, useful for treating diseases e.g., cancer or
PT neurodegenerative diseases.
PS Claim 1; SEQ ID NO 325; 99pp; English.
XX
XX The invention relates to a polynucleotide comprising a sequence given in
CC the specification, or its mature protein-coding portion, or its
CC complement. The polynucleotide is useful for treating diseases e.g.,
CC cancer or neurodegenerative diseases and many others listed in the
CC specification. The present sequence represents a novel human cDNA. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=2003104529.
XX
SQ Sequence 6475 BP; 1760 A; 1528 C; 1822 G; 1365 T; 0 U; 0 Other;
XX
Query Match 3.9%; Score 60; DB 9; Length 6475;
Best Local Similarity 61.5%; Pred. No. 0.0044;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGGAGCAGTGTCCGCTGTGTCGGGAGCGGGGAGCTCATCTGTCAGGCTGCCCTC 591
DB 1327 AGGACTATTGCGAGGTGTCCGCAAGGCGGTGAGATCTCTGTGTATCCTGTGCC 1386
QY 592 GGGCTTCACCTGGCTGCTGCTGCTCCCTCCGCTCCGGGAGATCCCGAGTGGGAGCTTGA 651
DB 1387 GTGCTTACCACTGCTGCTGCTGCTGATCCGACATGAGAAAGGCTCCCGAGGCAATGGA 1446
QY 652 GGTGCTCCAGCTGCTGCTGCAAGGCAACAGTCCAGAGG 687
DB 1447 GCTGCCACACTGCGAGAAAGGATCCAGTGGG 1482
XX
RESULT 21
ID AAI60230 standard; cDNA; 6553 BP.
XX
AC AAI60230;
XX
DT 22-OCT-2001 (first entry)
XX
DB Human polynucleotide SEQ ID NO 4219.
XX
XX Human; nocrotic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KM Leukemia; ss.
XX Homo sapiens.
OS WO200153312-A1.
XX
PN XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00639450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41074.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
PS Claim 1; SEQ ID NO 4219; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nocrotic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression.
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 6553 BP; 1802 A; 1538 C; 1829 G; 1384 T; 0 U; 0 Other;
XX
Query Match 3.9%; Score 60; DB 4; Length 6553;
Best Local Similarity 61.5%; Pred. No. 0.0044;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGGAGCAGTGTCCGCTGTGTCGGGAGCGGGGAGCTCATCTGCTGACGCTGCCCTC 591
DB 1304 AGGACTATTGCGAGGTGTCCGCAAGGCGGTGAGATCATCTGTGTATCCTGTGCC 1363
QY 592 GGGCTTCACCTGGCTGCTGCTGCTCCCTCCGCTCCGGGAGATCCCGAGTGGGAGCTTGA 651
DB 1364 GTGCTTACCACTGCTGCTGCTGATCCGACATGAGAAAGGCTCCCGAGGCAATGGA 1423
QY 652 GGTGCTCCAGCTGCTGCAAGGCAACAGTCCAGAGG 687
DB 1424 GCTGCCACACTGCGAGAAAGGATCCAGTGGG 1459
XX
RESULT 22
ID ADQ22446 standard; DNA; 6848 BP.
XX

DE cDNA encoding human Mi-2 antigen.
XX
XX Human; Mi-2 antigen; dermatomyositis; polymyositis;
KW myositis-sclerosis overlap disorder; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1053
FT /tag= a
FT /partial
FT /product= "Mi-2 antigen"
FT /note= "This sequence lacks both start and stop codons"
XX
XX US6610823-B1.
XX
XX
XX PD 26-AUG-2003.
XX
XX PF 09-SEP-1992; 92US-00945295.
XX
XX PR 07-SEP-1990; 90US-00579023.
XX
XX (UYOK-) UNITV OKLAHOMA.
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX
XX Targoff IN, Ge Q;
XX
XX WPI, 2003-895343/82.
XX P-PSDB; ADF69744.
XX
XX New human Mi-2 or PM-Scl protein, useful in assays for diagnosing
PT dermatomyositis, polymyositis or myositis-sclerosis overlap disorder.
XX
XX Claim 3; SEQ ID NO 1; 26pp; English.
XX
XX The present invention relates to the isolation of human antigens Mi-2 and
CC PM-Scl, and the polynucleotide sequences encoding them. The human Mi-2 or
CC PM-Scl proteins or a portion (at least one epitope) of them may be used
CC in assays for diagnosing dermatomyositis, polymyositis or myositis-
CC sclerosis overlap disorders. The present sequence encodes human Mi-2
CC antigen.
XX
XX Sequence 1053 BP; 265 A; 257 C; 340 G; 191 T; 0 U; 0 Other;
SQ
Query Match 3.9%; Score 59.8; DB 10; Length 1053;
Best Local Similarity 58.9%; Pred. No. 0.0036;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTCCGTCGTGTCGGAACGGGAGGAGCTCATCTGTGACGGCTGCCCTC 591
DB 248 AGGATTACTGTAGAGTGTGCCAGGAGGTGGGAATTATTCGTGTGACACTGCCCTC 307
QY 592 GGGCTTCCACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGAGACTTGA 651
DB 308 GTGGCTACCACTCGTCTGCTGATCTGAGCTTGAACGGGCTCCAGAGGGCAATGGA 367
QY 652 GGTGCTCCACGTCCTGCGGCAACAGTCCAGAGAGTGCAGCCCGGGGAGAGGA 706
DB 368 GCTGCCCTCACTGTGAGAGAGAGGGGCTCAGTGGAGGCCCAAGAGGAAGA 422

RESULT 25
ABV23250
ID ABV23250 standard; cDNA; 6331 BP.
XX
XX ABV23250;
XX
XX 16-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 23241.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.

XX
XX Homo sapiens.
OS
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX PD 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
XX 16-MAR-2000; 2000US-0189862P.
XX 25-MAY-2000; 2000US-0207454P.
XX 09-JUN-2000; 2000US-021314P.
XX 18-JUL-2000; 2000US-0219007P.
XX 13-DEC-2000; 2000US-0235281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endage WO, Monahan JB;
XX
XX WPI, 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 4195-4196; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 6331 BP; 1719 A; 1551 C; 1817 G; 1233 T; 0 U; 11 Other;
SQ
Query Match 3.9%; Score 59.8; DB 5; Length 6331;
Best Local Similarity 58.9%; Pred. No. 0.0048;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTCCGTCGTGTCGGAACGGGAGGAGCTCATCTGTGACGGCTGCCCTC 591
DB 1346 AGGATTACTGTAGAGTGTGCCAGGAGGTGGGAATTATTCGTGTGACACTGCCCTC 1405
QY 592 GGGCTTCCACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGAGACTTGA 651
DB 1406 GTGGCTACCACTCGTCTGCTGATCTGAGCTTGAACGGGCTCCAGAGGGCAATGGA 1465
QY 652 GGTGCTCCACGTCCTGCGGCAACAGTCCAGAGAGTGCAGCCCGGGGAGAGGA 706
DB 1466 GCTGCCCTCACTGTGAGAGAGAGGGGCTCAGTGGAGGCCCAAGAGGAAGA 1520

RESULT 26
ACD13374
ID ACD13374 standard; cDNA; 6331 BP.
XX
XX ACD13374;
XX
XX 13-AUG-2003 (first entry)
XX
XX Human DNA encoding a p53 modifier, SEQ ID 44.
XX
XX Human; ss; gene; p53 modifier; cytostatic; cancer; cytostatic;
KW antiangiogenic; antiapoptotic; p53 pathway; breast cancer; colon cancer;
KW kidney cancer; lung cancer; ovarian cancer; angiogenesis; cell cycle;

Query Match	3.9%	Score 59.8	DB 10	Length 6331
Best Local Similarity	58.9%	Pred. No. 0.0048		
Matches 103	Conservative 0	Mismatches 72	Indels 0	Gaps 0
Oy	5312	AGGAGAGAGTGTCCCTGTGTGTCCGAGACGGCGGGAGAGCTCATTCGTGTGTGACGGCTGCCTC	591	
Db	1346	AGGATTACTGTGTAGGTGTGTCCACGACGAGGGTGGGGAAATTATTCGTGTGTGACACTGCTC	140	
Oy	532	GGGCTTTCACCTGTGGCTGTGCTCTCCCTCCCTCCGCGAGATCCCACTGGACCTGGA	651	
Db	1406	GGCTTACCACTCTGTGTGCTCTGTATTCCTGACCTTGATCCGGGCTCCAGAGGGGCAATGGA	146	
Oy	652	GATGCTCCAGCTGCTCGCAGCAACAGTCAGAGGTGACAGCCCGGAGAGGA	706	
Db	1466	GCTGCTTCATCTGTGTGTGAGAGAGGGGGTTCAGTGGAGGCCAAGAGAGGAAGGA	1520	
RESULT 28				
ID	AD126768	standard; DNA; 6331 BP.		
XX	AD126768;			
AC	AD126768;			
XX	22-APR-2004 (first entry)			
DT	22-APR-2004 (first entry)			
XX				
DE	Human CHD3 DNA sequence, SEQ ID 29.			
XX				
KW	Cytostatic; ophthalmological; Antiarthritic; Antirheumatic;			
KW	Antiinflammatory; Antiartherosclerotic; Antipsoriatic; Gynaecological;			
KW	Vasotropic; Vulnereary; Hypotensive; Antianigmal; Cardiant; Osteopathic;			
KW	gene therapy; antiangiogenic; tumour vascularisation; retinopathy;			
KW	rheumatoid arthritis; Crohn disease; atherosclerosis;			
KW	ovarian hyperstimulation; psoriasis; endometriosis; neovascularisation;			
KW	restenosis; cardiovascular disease; hypertension; Raynaud disease;			
KW	muscular degeneration; osteoporosis; human; CHD3; gene; ds.			
XX				
OS	Homo sapiens.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	211..6045		
FT		/*tag= a		
FT		/product= "CHD3"		
XX				
FN	FR2837391-A1.			
XX				
PD	26-SEP-2003.			
XX				
PE	22-MAR-2002; 2002FR-00003655.			
XX				
PR	22-MAR-2002; 2002FR-00003655.			
XX				
PA	(GENE-) GENE SIGNAL.			
PA	(ALMA/) AL MAHMOOD S.			
XX				
PI	Colin S, Schneider C, Al Mahmood S;			
XX				
DR	WPI; 2003-806304/76.			
XX				
DR	P-PSDB; AD126772.			
XX				
PT	Antiangiogenic composition, useful for treating e.g. tumors, comprises			
PT	new angiogenesis-related nucleic acids or related polypeptides,			
PT	antibodies or antisense sequences.			
XX				
PS	Claim 2; SEQ ID NO 29; 110pp; French.			
XX				
CC	The present invention relates to antiangiogenic pharmaceutical			
CC	compositions. The compositions comprise, as active ingredient, one or			
CC	more of: nucleic acid (I; AD126740-AD126744 and AD126766-AD126769) from			
CC	an endothelial cell gene the expression of which is induced by an			
CC	angiogenic agent, or its complement or fragment; polypeptide (II;			
CC	AD126745-AD126749 and AD126770-AD126773) encoded by (I), or its fragment;			
CC	an antisense nucleic acid (III; AD126750-AD126753) that binds (II). The compositions			
CC	expression of (I); or an antibody (Ab) that binds (II). The compositions			

CC	are used for treating a wide range of angiogenic diseases: tumor
CC	vascularisation; retinopathy; rheumatoid arthritis; croum disease;
CC	atherosclerosis; ovarian hyperstimulation; psoriasis; endometriosis
CC	associated with neovascularization; restenosis (arterial or after balloon
CC	angioplasty); overgrowth of cells in wound healing; peripheral vascular
CC	diseases; hypertension; vascular inflammation; Raynaud disease; aneurysm;
CC	thrombophilicities; lymphedema; ischaemia; angina; myocardial infarct;
CC	chronic heart disease; (congestive) cardiac insufficiency; age-related
CC	muscular degeneration and osteoporosis. This sequence encodes CHD3.
CC	
XX	
XX	Sequence 6331 BP; 1720 A; 1552 C; 1819 G; 1240 T; 0 U; 0 Other;
XX	
XX	Query Match 3.9%; Score 59.8; DB 10; Length 6331;
XX	Best local similarity 58.9%; Pred. No. 0.0048;
XX	Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
Qy	532 AGACGAGTGTCCGCTGTGTGCGGAGCGCGGAGCTCATCTGTGACGGCTGCCCTC 591
Db	1346 AGGATTAAGTGTGAGAGGTGTGCCAGCAGGAGGGGGAATATTTCTGTGTACACCTGCCCTC 1405
Qy	592 GGGCTTCACACTGCGCCCTGTGCTGCCCTCCGCTCCGGGAGATCCCGAGTGGGACTGTGA 651
Db	1406 GTGCTTACACACTGTGCTGTGCTTGAATCTGAGCTTGACCGGCTCTCAAGAGGCAATGGA 1465
Qy	652 GATGCTTCAGCTGCGCTGTGAGGCAACAGTCCAGAGGTGACAGCCCGGCGCAGAGA 706
Db	1466 GCTGCGCTCACTGTGTGAGAGAGAGGGGGGTCCAGTGTGGAGGCCAAGAGGAAAGAA 1520
RESULT 29	
AD54629	
ID	AD54629 standard; DNA; 6331 BP.
XX	
AC	AD54629;
XX	
DT	26-JUN-2003 (first entry)
DE	
XX	Human chromodomain helicase DNA binding protein (CHD) encoding DNA #4.
XX	
KW	Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW	chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KW	cell proliferative disorder; chromatin organisation modifier domain;
KW	cytosolic; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200298899-A2.
XX	
PD	12-DEC-2002.
XX	
PF	03-JUN-2002; 2002WO-US017466.
XX	
XX	
PR	05-JUN-2001; 2001US-0296076P.
PR	10-OCT-2001; 2001US-0328605P.
PR	22-OCT-2001; 2001US-0338733P.
PR	15-FEB-2002; 2002US-0357253P.
PR	15-FEB-2002; 2002US-0357600P.
XX	
PA	(EXEL-) EXELIXIS INC.
PI	Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI	Lionblin M;
XX	
DR	WPI; 2003-156840/15.
XX	
PT	Identifying a candidate p53 pathway-modulating agent as therapeutic
PT	targets for disorders related to defective p53 function e.g. cancer by
PT	contracting an assay system having purified CHD polypeptide or nucleic
PT	acid, with a test agent.
XX	
PS	Disclosure; Page 51-54; 278pp; English.
CC	The present invention relates to a method for identifying candidate p53

CC pathway modulating agents. The method involves contacting an assay system
CC comprising purified chromatin organisation modifier (chromo) domain
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally
CC active fragments or derivatives, with a test agent under conditions
CC where, but for the presence of the test agent, the system provides a
CC reference activity. The methods are useful for identifying modulators of
CC the p53 pathway as therapeutic targets for disorders associated with
CC defective p53 function, such as angiogenic disorders, apoptotic disorders
CC or cell proliferative disorders, e.g. cancer. The modulators are useful
CC as research reagents, diagnostics and therapeutics. The invention is also
CC useful in gene therapy. The present sequence is human CHD DNA
XX
SQ Sequence 6331 BP; 1720 A; 1552 C; 1819 G; 1240 T; 0 U; 0 Other;
Query Match 3.9%; Score 59.8; DB 10; Length 6331;
Best Local Similarity 58.9%; Pred. No. 0.0049;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTCCGCTGTCGAGGAGGAGGAGCTCATCTGTCGACGCTGCCCTC 591
DB 1346 AGGATTACTGTAGAGGTGTCACAGAGGAGGAGAAATTATCTGTGTGACACTGCCCCC 1405
QY 592 GGGCCCTTCCACCTGAGCTGCTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGAGCTGGA 651
DB 1406 GTGCCCTACCACTCGTCTGCTGATCTGAGCTTGAACCGGCTCCAGAGGCAATGGA 1465
QY 652 GGTGCTCCAGCTGCTGTCAGGCAACATCCAGAGGTGACGCCCCGGGAGAGA 706
DB 1466 GCTGCCCTCACTGTGAGAGAGAGGAGGAGTCCAGTGGAGCCCAAGAGAAAGA 1520
RESULT 30
AD126742
ID AD126742 standard; DNA; 6771 BP.
XX
AC AD126742;
DT 22-APR-2004 (first entry)
XX
DE Human GS-N3 DNA sequence, SEQ ID 2.
XX
CYCOTACTIC; ophthalmological; Antiarthritic; Antirheumatic;
XX Antinflammatory; Antiartherosclerotic; Antisporadic; Gynaecological;
XX Vasectopic; Vulnereary; Hypocensive; Antianginal; Cardiant; Osteopathic;
XX gene therapy; antiangiogenic; tumour vascularisation; retinopathy;
XX rheumatoid arthritis; Crohn disease; atherosclerosis;
XX ovarian hyperstimulation; psoriasis; endometriosis; neovascularization;
XX restenosis; cardiovascular disease; hypertension; Raynaud disease;
XX muscular degeneration; osteoporosis; human; zing finger; helicase; gene;
XX de.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT CDS 151..6153
FT /tag= a
FT /product= "GS-P3"
XX
XX FR2837391-A1.
XX PD 26-SBP-2003.
XX
XX 22-MAR-2002; 2002FR-00003655.
XX PF 22-MAR-2002; 2002FR-00003655.
XX PR 22-MAR-2002; 2002FR-00003655.
XX
XX (GENE-) GENE SIGNAL.
XX PA (ALMA/) AL MAHMOOD S.
XX
XX Colin S, Schneider C, Al Mahmood S;
PI WPI; 2003-806304/76.
XX P-PSDB; AD126747.
DR

XX
PT Antiangiogenic composition, useful for treating e.g. tumors, comprises
PT new angiogenesis-related nucleic acids or related polypeptides,
PT antibodies or antisense sequences.
XX
PS Claim 2; SEQ ID NO 3; 110bp; French.
XX
CC The present invention relates to antiangiogenic pharmaceutical
CC compositions. The compositions comprise, as active ingredient, one or
CC more of: nucleic acid (I); AD126740-AD126744 and AD126766-AD126769) from
CC an endothelial cell gene the expression of which is induced by an
CC angiogenic agent, or its complement or fragment; polypeptide (II);
CC AD126745-AD126749 and AD126770-AD126773) encoded by (I), or its fragment;
CC an antisense nucleic acid (III; AD126750-AD126753) able to inhibit
CC expression of (I); or an antibody (Ab) that binds (II). The compositions
CC are used for treating a wide range of angiogenic diseases: tumour
CC vascularisation; retinopathy; rheumatoid arthritis; Crohn disease;
CC atherosclerosis; ovarian hyperstimulation; psoriasis; endometriosis
CC associated with neovascularization; restenosis (arterial or after balloon
CC angioplasty); overgrowth of cells in wound healing; peripheral vascular
CC disease; hypertension; vascular inflammation; Raynaud disease; aneurysm;
CC thrombophlebitis; lymphedema; ischaemia; angina; myocardial infarct;
CC chronic heart disease; (congestive) cardiac insufficiency; age-related
CC muscular degeneration and osteoporosis. GS-N3 encodes a zing finger
CC helicase protein.
XX
SQ Sequence 6771 BP; 1759 A; 1706 C; 2033 G; 1273 T; 0 U; 0 Other;
Query Match 3.9%; Score 59.8; DB 10; Length 6771;
Best Local Similarity 58.9%; Pred. No. 0.0049;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTCCGCTGTCGAGGAGGAGGAGCTCATCTGTCGACGCTGCCCTC 591
DB 1286 AGGATTACTGTAGAGGTGTCACAGAGGAGGAGAAATTATCTGTGTGACACTGCCCCC 1345
QY 592 GGGCCCTTCCACCTGAGCTGCTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGAGCTGGA 651
DB 1346 GTGCCCTACCACTCGTCTGCTGATCTGAGACTTGAACCGGCTCCAGAGGCAATGGA 1405
QY 652 GGTGCTCCAGCTGCTGTCAGGCAACATCCAGAGGTGACGCCCCGGGAGAGA 706
DB 1406 GCTGCCCTCACTGTGAGAGAGAGGAGGAGTCCAGTGGAGCCCAAGAGAAAGA 1460
RESULT 31
ABX63370
ID ABX63370 standard; cDNA; 7805 BP.
XX
AC ABX63370;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human cDNA #370 differentially expressed in activated vascular tissue.
XX
XX
XX Human; gene; ss; vascular tissue; cytosolic; atherosclerosis; cardiant;
XX hypotensive; antidiabetic; gynaecological; vasectopic; cerebroprotective;
XX gene therapy; vascular disease; cancer; coronary; artery disease;
XX hypertension; diabetes; pre-eclampsia; restenosis;
XX ischaemia-reperfusion injury; stroke.
XX
XX
XX Homo sapiens.
XX
XX US2002137081-A1.
XX PN US2002137081-A1.
XX
XX 26-SBP-2002.
XX PD 26-SBP-2002.
XX
XX 08-JAN-2002; 2002US-00044090.
XX PF 08-JAN-2002; 2002US-00044090.
XX PR 28-JUL-2000; 2000US-0222469P.
XX PR 08-JAN-2001; 2001US-0260483P.
XX
XX (BAND/) BANDMAN O.
PA

XX Bandman O;
PI
XX
DR WPI; 2003-110597/10.
XX
PT Combination for diagnosing, staging, treating, or monitoring the
PT progression of treatment of a vascular disease, e.g. atherosclerosis,
PT comprises several cDNAs that are differentially expressed in activated
PT vascular tissue.
XX
PS Claim 1; Page; 18pp; English.
XX
XX This invention relates to a combination comprising several cDNAs that are
CC differentially expressed in activated vascular tissue. The invention also
CC discloses a high throughput method for detecting differentially expressed
CC cDNAs in a sample. The cDNAs of the invention may have
CC antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC gynaecological; vasotropic and cerebroprotective activities and may be
CC used in gene therapy. The cDNAs of the invention may be used in a high-
CC throughput methods for detecting differential expression of one or more
CC cDNAs in a sample, or screening several molecules or compounds to
CC identify a molecule or compound that specifically binds a cDNA of the
CC invention. A protein encoded by the cDNA may be used to screen several
CC molecules or compounds to identify a ligand that specifically binds to
CC the protein, or to produce or purify an antibody to the protein that can
CC be used to detect a protein in a sample or purify a natural or
CC recombinant protein from a sample. The nucleotides may be useful for
CC diagnosing, staging, treating, or monitoring the progression of treatment
CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion
CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale
CC genetic or gene expression analysis of several new nucleic acid
CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC diagnosing pre-pathologic disorders, and chronic or acute diseases
CC associated with abnormalities in the expression, amount or distribution
CC of the protein. The present sequence represents a cDNA of the invention
CC that is differentially expressed in activated vascular tissue. Note: The
CC sequence data for this patent did not form part of the specification, but
CC was obtained in electronic format directly from USPTO at
CC http://seqdata.uspto.gov/sequence.html?docID=20020137081
XX
SQ Sequence 7805 BP; 2003 A; 1945 C; 2323 G; 1534 T; 0 U; 0 Other;
XX
XX
Query Match 3.9%; Score 59.8; DB 8; Length 7805;
Best Local Similarity 58.9%; Pred. No. 0.005;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
XX
QY 532 AGACGAGTGTGCGGTCGCGGAGCGGCGGAGCTCATCTGCTGTGACGGCTGCGCTC 591
DB 1630 AGGATTACTGTAGGTGTCGACGAGGTGGGGAATTAATTCTGTGACACTGCGCTC 1689
QY 592 GGGCCCTTCACCTGGCTGCGCTGCTCCCTCCGTCGCGGAGATCCCAATGGGAACTGGA 651
DB 1690 GTGCTTACACCTCGTCTGCTGATCTGAGCTTGACCGGGCTCCAGAGGGGAATGGA 1749
QY 652 GGTGCTCCAGCTGCTGTCGACGCAAGTCAGAGAGGTGACGCCCGGCGAGAGA 706
DB 1750 GCTGCCCTCACTGTGAGAAAGAGGGGCTCACTGCGAGGCCCAAGAGGAAGA 1804
XX
RESULT 32
AD122590
ID AD122590 standard; DNA; 7805 BP.
XX
AC AD122590;
XX
DT 22-APR-2004 (first entry)
XX
DE Human liver differentially expressed cDNA seq id 400.
XX
KM hepatotropic; antiinflammatory; virucide; cytostatic; gene therapy;
KM differentially gene expression; liver; toxin; liver disorder;
KM biliary cirrhosis; X-linked adrenoleukodystrophy; Zellweger syndrome;

KW hepatorenal syndrome; hepatitis; hepatocarcinoma; metabolic response;
KM toxicological response; ss; EST; expressed sequence tag; human.
XX
OS Homo sapiens.
XX
PN US2003165854-A1.
XX
PD 04-SRP-2003.
XX
PF 05-DEC-2001; 2001US-00006285.
XX
PR 05-DEC-2000; 2000US-0251986P.
XX
PA (CUNN/) CUNNINGHAM M J.
PA (KASE/) KASER M R.
PI Cunningham MJ, Kaser MR,
PI
DR WPI; 2003-863697/80.
XX
PT New combination comprising a number of cDNAs that are differentially
PT expressed in a liver treated with a toxin, useful for diagnosing, staging
PT or treating liver disorders (e.g. biliary cirrhosis, hepatitis or
PT hepatocarcinoma).
XX
PS Claim 1; SEQ ID NO 400; 28pp; English.
XX
XX The invention describes a combination comprising a number of cDNAs that
CC are differentially expressed in a liver treated with a toxin and are
CC selected from any of the 514 cDNAs listed in the specification, or their
CC complements. The combination is useful in diagnosing, staging or treating
CC a liver disorder (e.g. biliary cirrhosis, X-linked adrenoleukodystrophy,
CC Zellweger syndrome, hepatorenal syndrome, hepatitis or hepatocarcinoma),
CC in monitoring diagnostic and therapeutic applications, in detecting
CC metabolic and toxicological responses, and in elucidating drug mechanism
CC of action. This sequence represents a cDNA differentially expressed in
CC liver tissues in response to treatment with a toxin.
XX
SQ Sequence 7805 BP; 2003 A; 1945 C; 2323 G; 1534 T; 0 U; 0 Other;
XX
XX
Query Match 3.9%; Score 59.8; DB 10; Length 7805;
Best Local Similarity 58.9%; Pred. No. 0.005;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
XX
QY 532 AGACGAGTGTGCGGTCGCGGAGCGGCGGAGCTCATCTGCTGTGACGGCTGCGCTC 591
DB 1630 AGGATTACTGTAGGTGTCGACGAGGTGGGGAATTAATTCTGTGACACTGCGCTC 1689
QY 592 GGGCCCTTCACCTGGCTGCGCTGCTCCCTCCGTCGCGGAGATCCCAATGGGAACTGGA 651
DB 1690 GTGCTTACACCTCGTCTGCTGATCTGAGCTTGACCGGGCTCCAGAGGGGAATGGA 1749
QY 652 GGTGCTCCAGCTGCTGTCGACGCAAGTCAGAGAGGTGACGCCCGGCGAGAGA 706
DB 1750 GCTGCCCTCACTGTGAGAAAGAGGGGCTCACTGCGAGGCCCAAGAGGAAGA 1804
XX
RESULT 33
ADA71938/C
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KM Plant; bacterial infection; fungal infection; viral infection; rice;
KM gene; ds.
XX
OS Oryza sativa.
OS
PN WO200300898-A1.

CC colon cancer, breast cancer, lung cancer, pancreatic cancer,
CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as
CC well as all types of cancers which may be metastasized or have metastasized
CC to the lungs, including breast and prostate cancer
XX
SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

Query Match 3.7%; Score 57.4; DB 2; Length 114955;

Best Local Similarity 29.7%; Pred. No. 0.023; Indels 5; Gaps 2;
Matches 301; Conservative 104; Mismatches 604;

QY 257 CGGGGCCCCCTGGAGACGACGACCTGCAAGAAACCGGGTTTCTTCCCAATAGGATGAC 316
DB 104081 CCGGGCCSNNNDNNCGCGTGGGCGGGGCSNNNDNNGGTGGGCGGGGCSNNNDNNCC 104140
QY 317 CCGGGGGGTGTCTGTTCGAGACAGAGTGGGAGACAGTGGTCTGAGGGCAATTTTC 376
DB 104141 GTGGGGCCGGGCSNNNDNNCGTGGGCGGGGCSNNNDNNGTGGGCGGGGCSNNNDNNTC 104200
QY 377 AGGCGCTGGACATGAGAGGAGGAGAGTGGGGAATTCAGATACCCAGAGATGCTG 436
DB 104201 GGGCGGGCSNNNDNNCGGGCGGCSNNNDNNCGCBGGCCBG--GCGGCGCGCG 104256
QY 437 CTGGGGGAGCTGTTTGGAGAGAGGTGCTCTAGAGAGGTGCTGACACCCAGCCAGT 496
DB 104257 GCGGGGCSNNNDNNCGCBGGCGGGGCGGGCGGGGCGGGGCSNNNDNNCCGCBG 104316
QY 497 CTGCAATGGGCGTCTCTTCTGCTGCTGCAAGAAATGAGAGAGTGTGCTGTGTGGGA 556
DB 104317 GCBGGGGCGCGCGCGCGGCGGCSNNNDNNCGCBGGCGCGCGCGCGCGCGG 104376
QY 557 GCGCGGGAGCTCATCTGTGTGAGCGCTGCGGCGCTTGCACCTGGCGTGGCTGTC 616
DB 104377 SNNNDNNCGGCGGCGCBGGCGCGCGCGCGGCSNNNDNNCGCBGGCGCGCG 104436
QY 617 CCTCTCGCTCCGAGAGATCCCAAGTGGAGACTGAGAGTCTCACTGCTGCTGAGCAAC 676
DB 104437 GCGGGGCGSNNNDNNCGCBGGCGGGGCGGGCGCGGCSNNNDNNCGCBGGCGG 104496
QY 677 AGTCCAGAGAGTGCACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 736
DB 104497 GCGCGGCGCGCGGCSNNNDNNCGCBGGCGCGCGCGCGGCSNNNDNNCGCBGGC 104556
QY 737 CCGGCTCCCGCGGGCTTAAAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 796
DB 104557 CCGGGGCGCGCGGCSNNNDNNCGCBGGCGGGGCGGGCGCGGCSNNNDNNCGCBGGC 104616
QY 797 CCTAGCGGAGATGAGACAGACTCTTGTCTTACAGAGAGAGAGAGAGAGAGAGAGAG 856
DB 104617 BGGGGGCGCGGCSNNNDNNCGCBGGCGGGGCGGGCGGCSNNNDNNCGCBGGCGG 104676
QY 857 CCGGCTGCGAGGTGTGAGATCTCTCGGCGCTGACACCCCTTACTGTGTGTGGGTCT 916
DB 104677 GCGGCSNNNDNNCGCBGGCGGGGCGGCSNNNDNNCGCBGGCGGCGGCSNNNDN 104736
QY 917 TCAGCAGAGACTGGCTCTGTGTGCGGCTTGGCGGAGTGTATGAGAGAGAGAGAG 976
DB 104737 NCCGCBGGCGCGGCSNNNDNNCGCBGGCGGGGCSNNNDNNCGCBGGCGGGG 104796
QY 977 GCGGTGTACTCACTGCGCGCTGCTTCACTGCGCTGCACTTCCAGCGCGGCACTTC 1036
DB 104797 SNNNDNNCGCBGGCGGCSNNNDNNCGCBGGCGGGGCSNNNDNNCGCBGGCGG 104856
QY 1037 CCGGCGCGGAG 1096
DB 104857 NDNNDNNCGCBGGCGGCSNNNDNNCGCBGGCGGGGCGGGGCGGCGCGCG 104915
QY 1097 GAG 1156
DB 104916 GCGGGGCGSNNNDNNCGCBGGCGGGGCGGGCGGGGCGGCSNNNDNNCGBGGC 104975
QY 1157 TGCAGTCAAGAGCGGCTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1216

DB 104976 GGGGCGGCGCGCGCGGCSNNNDNNBGGCCBGGGCGGCGCGCGGCGGCGSNN 105035
QY 1217 CTTGATGAGCATCTCTGAGTGGGCGCATCCAGAGATGAGCGGCGGGGCGGCCCC 1270
DB 105036 NDNNDNNCGCBGGCGGCGCGCGGCGGCGGCSNNNDNNCGCBGGGCGGCGGCG 105089

RESULT 35

AB109061
ID AB109061 standard; cDNA; 6124 BP.

AB109061;

26-MAR-2002 (first entry)

Drosophila melanogaster expressed polynucleotide SEQ ID NO 21655.

Drosophila; developmental biology; cell signalling; insecticide;

pharmaceutical; gene; ss.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US009231.

23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

PA (PEKE) PE CORP NY.

Venter JC, Adams M, Li PMD, Myers EW;

WPI; 2001-65860/75.

P-PSDB; ABB64958.

New isolated nucleic acid detection reagent for detecting 1000 or more

genes from Drosophila and for elucidating cell signaling and cell-cell

interactions.

Claim 1; SEQ ID NO 21655; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signaling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
sequences (AB101840-AB116175) and the encoded proteins (AB57737-
AB572072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 6124 BP; 1706 A; 1562 C; 1509 G; 1347 T; 0 U; 0 Other;

Query Match 3.7%; Score 56.4; DB 4; Length 6124;

Best Local Similarity 59.3%; Pred. No. 0.025; Indels 0; Gaps 0;
Matches 96; Conservative 0; Mismatches 66;

QY 528 AATGAGAGAGAGTGTGCTGTGTGCGAGAGCGGAGAGTCTGCTGTGAGAGAGAG 587
DB 3391 AACGAGAGAGTGTGCGCTGTGTGTGATGAGAGAGAGAGTGTGCTGTGAGAGAG 3450

QY 588 CCTGGGCGCTTCACTGAGCTGCTGTGCTTCTGCTGCTGCGGAGATCCCAAGTGGAGC 647
DB 3451 CCCAAGTTTTCATCAAGATGTCAATCTCTGAGATCAAGTCTTGGCGAGCAGAGAC 3510

QY 648 TGAAGTGTCTCAGTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 689
DB 3511 GAGAGTGTCAAGTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3552

RESULT 36
ABLO9060
ID ABL09060 standard; cDNA; 21747 BP.
XX
AC ABL09060;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 21662.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-656660/75.
XX
DR P-PsDB; ABB64957.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 21662; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-
CC ABB12072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 21747 BP; 6181 A; 4697 C; 4637 G; 6232 T; 0 U; 0 Other;
Query Match 3.7%; Score 56.4; DB 4; Length 21747;
Best Local Similarity 59.3%; Pred. No. 0.03;
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 528 AATGAGAGAGATGTCCTGTCGCGAGCGCGGAGACTATCTGCTGTGACGCTTC 587
DB 17810 AACGAGAGATGTCGCGCTCTGTCTGATGAGGAGGAGACTATGTCGACAGATGT 17869
QY 588 CTTGCGGCTTCCACCTGCGCTGCTGCTCCCTCCGCTCCGAGATCCCACTGGGACC 647
DB 17870 CCCAAGTTTTCATCAACAGCTGATCATCTCCCTGCGATAGCTCGTTCGAGACGAGAC 17929
QY 648 TGGAGTGTCTCAGTGTCTGAGGAGACAGTCCAGAGAGTG 689
DB 17930 GAGAGCTGGCAGTGTCTACTGTGCGTCAATCAAGAGAGCTG 17971

RESULT 37
ACH90409/c
ID ACH90409 standard; DNA; 153 BP.
XX
AC ACH90409;
XX
DT 29-JUL-2004 (first entry)
XX
DE Human genome derived single exon probe #23604.
XX
KW Human; probe; ss; gene expression; single exon probe; microarray;
XX alternative splicing event; genomic alteration.
XX
OS Homo sapiens.
XX
PN US2003194704-A1.
XX
PD 16-OCT-2003.
XX
PF 03-APR-2002; 2002US-00029386.
XX
PR 03-APR-2002; 2002US-00029386.
XX
PA (PENN/) PENN S G.
XX
PA (RANK/) RANK D R.
XX
PA (HANZ/) HANZEL D K.
XX
PI Penn SG, Rank DR, Hanzel DK;
XX
DR WPI; 2004-119264/12.
XX
PT New human genome-derived single exon nucleic acid probes useful for human
PT gene expression analysis, for identifying or characterizing alternative
PT splicing events, for assessing genomic alterations or as tools for
PT surveying tissues.
XX
PS Claim 1; SEQ ID NO 23604; 80bp; English.
XX
CC The invention relates to a nucleic acid probe for measuring human gene
CC expression, comprising any of the 27,400 fully defined nucleotide
CC sequences in the specification, or their complements or fragments, and
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
CC fully defined in the specification. The probe is a single exon probe that
CC hybridises under high stringency conditions to a nucleic acid molecule
CC expressed in human cells or tissues. Also included are a spatially-
CC addressable set of single exon nucleic acid probes for measuring human
CC gene expression (comprising a plurality of single exon nucleic acid
CC probes cited above, where each of the plurality of probes is separately
CC and addressably isolatable or amplifiable from the plurality), a single
CC exon microarray for measuring human gene expression, a method of
CC measuring human gene expression, a vector comprising the single exon
CC probe cited above, an ORF-encoded peptide comprising at least 8
CC contiguous amino acids of any of the above-mentioned amino acid
CC sequences (optionally with conservative amino acid substitutions), an
CC isolated antibody that binds specifically to a peptide cited above,
CC methods of selling and/or licensing single exon probes or microarrays to
CC a customer desiring to measure gene expression, a method of providing
CC human gene expression data by subscription, and a computer-readable
CC storage medium which contains a database having a plurality of records
CC (each record including data on the expression of a single exon probe
CC cited above. The probe, methods and apparatus are useful in gene
CC expression analysis. The probes may be used as tools for surveying
CC tissues to detect the presence of expressed messages that contain their
CC specific exon, or in constructing genome-derived single exon microarrays.
CC In addition, the probes are used in identifying and characterizing
CC alternative splicing events, in detecting and characterizing gross
CC alterations in the genomic locus that includes their exon, in assessing
CC smaller genomic alterations, in priming the synthesis of nucleic acids,
CC or in expressing the ORF-encoded peptide. The present sequence is a human
CC single exon probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC segdata.uspto.gov/sequence.html?docID=20030194704
XX
SQ Sequence 153 BP; 29 A; 48 C; 43 G; 33 T; 0 U; 0 Other;
Query Match 3.6%; Score 55.4; DB 12; Length 153;
Best Local Similarity 62.8%; Pred. No. 0.025;

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 2495 BP; 696 A; 581 C; 733 G; 485 T; 0 U; 0 Other;
XX
XX Query Match 3.6%; Score 55.4; DB 5; Length 2495;
XX Best Local Similarity 62.8%; Pred. No. 0.037;
XX Matches 86; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGACGCGGAGCTCATCTGTGACGGCTGCTTC 591
DB 1367 AGGACCTATTGCGAGGTGTGCGGACGCGGAGCTCATCTGTGATTAAGCTGCTCC 1426
QY 592 GGGCTTCCACCTGCGCTGCTGCTCCCTCCGCTCCGGAAGATCCCAAGTGGAGCTTGA 651
DB 1427 GTGCTTACCAATGCTGCTGCTGATCCGACATGAGAAAGCTCCCGAGGCAAGTGA 1486
QY 652 GGTGCTCCAGCTGCTG 668
DB 1487 GGTGCCACACCTGCTG 1503

RESULT 40
AA574874
ID AA574874 standard; cDNA; 7132 BP.
XX
XX AA574874;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #10678.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dermanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG10687.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 10678; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 7132 BP; 1723 A; 1875 C; 1964 G; 1569 T; 0 U; 1 Other;
XX
XX Query Match 3.6%; Score 55.4; DB 5; Length 7132;
XX Best Local Similarity 62.8%; Pred. No. 0.042;
XX Matches 86; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGACGCGGAGCTCATCTGTGACGGCTGCTTC 591
DB 1367 AGGACCTATTGCGAGGTGTGCGGACGCGGAGCTCATCTGTGATTAAGCTGCTCC 1426
QY 592 GGGCTTCCACCTGCGCTGCTGCTCCCTCCGCTCCGGAAGATCCCAAGTGGAGCTTGA 651
DB 1427 GTGCTTACCAATGCTGCTGCTGATCCGACATGAGAAAGCTCCCGAGGCAAGTGA 1486
QY 652 GGTGCTCCAGCTGCTG 668
DB 1487 GGTGCCACACCTGCTG 1503

RESULT 41
AAK71358
ID AAK71358 standard; DNA; 37314 BP.
XX
XX AAK71358;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26170.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytotoxic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX

PR 17-MAR-2000; 2000US-0190076P-
PR 18-APR-2000; 2000US-0198123P-
PR 19-MAY-2000; 2000US-0205515P-
PR 07-JUN-2000; 2000US-0209467P-
PR 28-JUN-2000; 2000US-0214886P-
PR 30-JUN-2000; 2000US-0215135P-
PR 07-JUL-2000; 2000US-0216647P-
PR 07-JUL-2000; 2000US-0216880P-
PR 11-JUL-2000; 2000US-0217487P-
PR 11-JUL-2000; 2000US-0217496P-
PR 14-JUL-2000; 2000US-0218290P-
PR 26-JUL-2000; 2000US-0220963P-
PR 14-AUG-2000; 2000US-0224518P-
PR 14-AUG-2000; 2000US-0224519P-
PR 14-AUG-2000; 2000US-0225213P-
PR 14-AUG-2000; 2000US-0225214P-
PR 14-AUG-2000; 2000US-0225266P-
PR 14-AUG-2000; 2000US-0225267P-
PR 14-AUG-2000; 2000US-0225268P-
PR 14-AUG-2000; 2000US-0225270P-
PR 14-AUG-2000; 2000US-0225447P-
PR 14-AUG-2000; 2000US-0225757P-
PR 14-AUG-2000; 2000US-0225758P-
PR 14-AUG-2000; 2000US-0225759P-
PR 18-AUG-2000; 2000US-0226279P-
PR 22-AUG-2000; 2000US-0226681P-
PR 22-AUG-2000; 2000US-0226686P-
PR 22-AUG-2000; 2000US-0227182P-
PR 23-AUG-2000; 2000US-0227009P-
PR 30-AUG-2000; 2000US-0228924P-
PR 01-SEP-2000; 2000US-0229287P-
PR 01-SEP-2000; 2000US-0229343P-
PR 01-SEP-2000; 2000US-0229344P-
PR 01-SEP-2000; 2000US-0229345P-
PR 05-SEP-2000; 2000US-0229509P-
PR 06-SEP-2000; 2000US-0230437P-
PR 06-SEP-2000; 2000US-0230438P-
PR 08-SEP-2000; 2000US-0231242P-
PR 08-SEP-2000; 2000US-0231243P-
PR 08-SEP-2000; 2000US-0231244P-
PR 08-SEP-2000; 2000US-0231413P-
PR 08-SEP-2000; 2000US-0231414P-
PR 08-SEP-2000; 2000US-0232080P-
PR 08-SEP-2000; 2000US-0232081P-
PR 12-SEP-2000; 2000US-0231968P-
PR 14-SEP-2000; 2000US-0232397P-
PR 14-SEP-2000; 2000US-0232398P-
PR 14-SEP-2000; 2000US-0232399P-
PR 14-SEP-2000; 2000US-0232400P-
PR 14-SEP-2000; 2000US-0232401P-
PR 14-SEP-2000; 2000US-0233063P-
PR 14-SEP-2000; 2000US-0233064P-
PR 14-SEP-2000; 2000US-0233065P-
PR 21-SEP-2000; 2000US-0234223P-
PR 21-SEP-2000; 2000US-0234274P-
PR 25-SEP-2000; 2000US-0234897P-
PR 25-SEP-2000; 2000US-0234988P-
PR 26-SEP-2000; 2000US-0235484P-
PR 27-SEP-2000; 2000US-0235834P-
PR 27-SEP-2000; 2000US-0235836P-
PR 29-SEP-2000; 2000US-0236327P-
PR 29-SEP-2000; 2000US-0236367P-
PR 29-SEP-2000; 2000US-0236368P-
PR 29-SEP-2000; 2000US-0236369P-
PR 29-SEP-2000; 2000US-0236370P-
PR 02-OCT-2000; 2000US-0236802P-
PR 02-OCT-2000; 2000US-0237037P-
PR 02-OCT-2000; 2000US-0237038P-
PR 02-OCT-2000; 2000US-0237039P-
PR 02-OCT-2000; 2000US-0237040P-
PR 13-OCT-2000; 2000US-0239935P-

PR 13-OCT-2000; 2000US-0239937P-
PR 20-OCT-2000; 2000US-0240960P-
PR 20-OCT-2000; 2000US-0241221P-
PR 20-OCT-2000; 2000US-0241785P-
PR 20-OCT-2000; 2000US-0241786P-
PR 20-OCT-2000; 2000US-0241787P-
PR 20-OCT-2000; 2000US-0241808P-
PR 20-OCT-2000; 2000US-0241809P-
PR 20-OCT-2000; 2000US-0241826P-
PR 01-NOV-2000; 2000US-0244617P-
PR 08-NOV-2000; 2000US-0246474P-
PR 08-NOV-2000; 2000US-0246475P-
PR 08-NOV-2000; 2000US-0246525P-
PR 08-NOV-2000; 2000US-0246526P-
PR 08-NOV-2000; 2000US-0246527P-
PR 08-NOV-2000; 2000US-0246528P-
PR 08-NOV-2000; 2000US-0246532P-
PR 08-NOV-2000; 2000US-0246609P-
PR 08-NOV-2000; 2000US-0246610P-
PR 08-NOV-2000; 2000US-0246611P-
PR 08-NOV-2000; 2000US-0246613P-
PR 17-NOV-2000; 2000US-0249207P-
PR 17-NOV-2000; 2000US-0249208P-
PR 17-NOV-2000; 2000US-0249209P-
PR 17-NOV-2000; 2000US-0249210P-
PR 17-NOV-2000; 2000US-0249211P-
PR 17-NOV-2000; 2000US-0249212P-
PR 17-NOV-2000; 2000US-0249213P-
PR 17-NOV-2000; 2000US-0249214P-
PR 17-NOV-2000; 2000US-0249215P-
PR 17-NOV-2000; 2000US-0249216P-
PR 17-NOV-2000; 2000US-0249217P-
PR 17-NOV-2000; 2000US-0249218P-
PR 17-NOV-2000; 2000US-0249244P-
PR 17-NOV-2000; 2000US-0249245P-
PR 17-NOV-2000; 2000US-0249264P-
PR 17-NOV-2000; 2000US-0249265P-
PR 17-NOV-2000; 2000US-0249297P-
PR 17-NOV-2000; 2000US-0249299P-
PR 17-NOV-2000; 2000US-0249300P-
PR 01-DEC-2000; 2000US-0250160P-
PR 01-DEC-2000; 2000US-0250391P-
PR 05-DEC-2000; 2000US-0251030P-
PR 05-DEC-2000; 2000US-0251988P-
PR 05-DEC-2000; 2000US-0256719P-
PR 06-DEC-2000; 2000US-0251479P-
PR 08-DEC-2000; 2000US-0251856P-
PR 08-DEC-2000; 2000US-0251868P-
PR 08-DEC-2000; 2000US-0251869P-
PR 08-DEC-2000; 2000US-0251989P-
PR 08-DEC-2000; 2000US-0251990P-
PR 11-DEC-2000; 2000US-0254097P-
PR 05-JAN-2001; 2001US-0259678P-

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Barash SC, Ruben SM,
XX
XX WPI; 2001-483426/52.
DR
XX
XX Nucleic acids encoding human immune/nematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX
PS Disclosure; SEQ ID NO 26170; 3071pp + Sequence Listing; English.
CC
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM62170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)

PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Metberg D, Treacy M, Bowman MR;
XX WPI: 2000-317938/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (ESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders.
XX
XX Claim 1, Page 695; 803pp; English.
XX
XX AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (ESTs), isolated from human, mouse, chicken and rat tissue
CC sources. The ESTs can have a range of activities depending on the
CC tissues they were isolated from. The activities include: chemotactic;
CC proliferative; immunomodulatory; haematopoietic; chemokine; analgesic;
CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
CC antifungal; antiviral; antidiabetic; antiallergic; antitumor; anticancer;
CC osteoprotective; neuroprotective; neurotropic; antiparkinsonian; antiproliferative;
CC cerebroprotective; anticonvulsant; and antidepressant. The ESTs can be
CC used for gene therapy and in vaccines. The ESTs are useful as probes for
CC the identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the ESTs. Proteins encoded by the ESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma, myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
SQ Sequence 523 BP; 134 A; 113 C; 167 G; 96 T; 0 U; 13 Other;
Query Match 3.6%; Score 55.2; DB 3; Length 523;
Best Local Similarity 59.6%; Pred. No. 0.033;
Matches 93; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTGCTGCTGCGGACGCGGGAGCTCATCTGCTGACGCGCTCC 591
DB 186 AGGACTATTGCGAGGTGTGCGACCAAGGGGAGAGATCATCTGTGTATCCTGTCC 245
QY 592 GGGGCTTCACCGCTGCTGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGAGCTGGA 651
DB 246 GAGCTTACCAATGATGATGCTGCTGAGACCAAGATGAGAGAGGCCCGAGGGCAAGTGA 305
QY 652 GGTGTCTCAGCTGCTGCGAGGCAACAGTCCAGAGG 687
DB 306 GCTGTCCCACTGTGAGAGAGAGGAGATCCAGTGGG 341
RESULT 44
ID ACH48320 standard; cDNA; 494 BP.
XX ACH48320;
AC
XX
DT 13-OCT-2003 (first entry)
XX
DE Human lung tumour CDNA #453.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KM genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.

XX
PR 30-JUL-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI: 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1, SEQ ID NO 35532; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC segdata.uspto.gov/sequence.html?docID=20030073623
XX
SQ Sequence 494 BP; 154 A; 101 C; 102 G; 136 T; 0 U; 1 Other;
Query Match 3.5%; Score 54.2; DB 9; Length 494;
Best Local Similarity 61.9%; Pred. No. 0.053;
Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 528 AATGAGGAGAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGCTGACGCGCTGC 587
DB 109 AATGAGGAGTGTGTGCACTTGTGTCAAAACGAGGGAACCTCTGCTGTGAAAAATGC 168
QY 588 CCTCGGACCTTCACACTGCTGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGAGCC 647
DB 169 CCCAAGATTTTCATCTTTCTGTGATGAGCCACATTTGACAAATTTTCAAGTGGAGAG 228
QY 648 TGGAGGTGCTCCAGCTGCC 666
DB 229 TGGATTTGCACTTCTGCC 247
RESULT 45
ID ADB80995 standard; DNA; 3039 BP.
XX ADB80995;
AC
XX
DT 04-DEC-2003 (first entry)
XX
DE RING-SH complex related DNA, SEQ ID No 69.
XX
XX RING-SH 3; Gag protein; Gag late domain; PI3K; actin; myosin; Hsp60;
KM Hsp70; Hsp90; STM1; STAM2; STRAM2; VHS-UIM; GTPase; E2 enzyme; tsq101;
KM cullin; RING-SH; clathrin; v-src; vaccine; antiviral; retrovirus;
KM rhadovirus; filovirus; gene; ds.
XX
OS unidentified.

XX		MO200303646-A2.	
PX			
XX		24-APR-2003.	
PD			
XX			
PF		31-JUL-2002; 2002WO-US024589.	
XX			
PR		31-JUL-2001; 2001US-0308958P.	
XX		09-NOV-2001; 2001US-0345846P.	
XX			
PA	(PROT-)	PROTEOLOGICS INC.	
XX			
PI	Greener T,	Moskowitz H, Reis Y, Alroy I;	
XX			
DR	WPI; 2003-393509/37.		
XX	P-P5DB; ADB80954.		
PT			
XX	New isolated protein complex comprising a RING-SH 3 polypeptide and another polypeptide, useful for detecting cells infected with a virus, or for treating viral disorders caused by retroviruses, rhadoviruses, or flaviviruses.		
PS	Disclosure; Fig 68; 176pp; English.		
XX			
CC	The invention relates to a novel isolated protein complex comprising a RING-SH 3 polypeptide and a novel polypeptide selected from a Gag protein, a Gag late domain, PI3K, actin, myosin, Hep60, Hep90, STAM1, STAM2A, STAM2B, VHS-UIM, a GTPase, an E2 enzyme, teg101, a cullin, RING-SH, and a clathrin. The novel protein complex has virucide activity and can be used to treat disorders as part of a vaccine. The protein complex and composition are useful for detecting cells infected with a virus, for identifying agents having antiviral activity, and for treating viral disorders caused by retroviruses, rhadoviruses, or flaviviruses. This polynucleotide represents a DNA sequence relating to a protein comprising the RING-SH complex of the invention.		
CC			
XX	Sequence 3039 BP; 941 A; 738 C; 664 G; 696 T; 0 U; 0 Other;		
SQ			
	Query Match	3.5%; Score 54.2; DB 9; Length 3039;	
	Best Local Similarity	61.9%; Pred. No. 0.068;	
	Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;		
OY	528 AATAGGACGAGTCGCCCTGTCTGGGAGACGCGGGAGCTCATCTGCTGTGACGCCTGC	587	
Dd	2359 AATGAGGACTG3GTGTGACAGTTTGTAATAACGAGGGGAACTCTCTGTGAATAAGTGC	2418	
OY	588 CCTCGGGGCTTCCACACTGAGCCTGCTGCTCCCTCGCTCCGAGAGATCCCAGTGGGACC	647	
Dd	2419 CCCAAGAATTCCACTCTTCTTGTGTATGTCGCCCACTTGACAAATTTTCCAAGTGGAGAG	2478	
OY	648 TGGAGTGTCTCCAGCTGCC	666	
Dd	2479 TGGATTGCACTTCTGCCC	2497	
	RESULT 46		
ID	ADJ56478 standard; CDNA; 3917 BP.		
XX	ADJ56478;		
XX	ADJ56478; (first entry)		
DT	06-MAY-2004		
DE	Human CDNA differentially expressed in MYCN activated cells Segid 284.		
XX	human; differential expression; transactivator; proto-oncogene; neuroblastoma; small cell lung cancer; cytostatic; gene therapy; ss;		
KM	MYCN activated cell.		
XX	Homo sapiens.		
OS			
XX	US2003119009-A1.		
FN			

PD	26-JUN-2003.
XX	
PF	25-FEB-2002; 2002US-00084817.
XX	
PR	23-FEB-2001; 2001US-0270784P.
XX	
PA	(STUVA/) STUART S G.
XX	
PA	(NUCH/) NUCHTERN J G.
XX	
PA	(PLON/) PLON S E.
XX	
PA	(SHOH/) SHOHEI J M.
XX	
PI	Stuart SG, Nuchtern JG, Plon SE, Shohet JM;
XX	
DR	WPI; 2003-635698/60.
XX	
PT	New genes regulated by MYCN activation, useful in gene therapy,
XX	
PT	particularly for treating a subject with e.g. neuroblastoma or other
XX	
PT	cancers, or for diagnosing, staging or monitoring the treatment of the
XX	
PT	cancer.
XX	
PS	Claim 1; SEQ ID NO 284; 27bp; English.
XX	
CC	This invention relates to novel isolated cDNAs that are differentially
XX	
CC	expressed in MYCN activated cells. Specifically, it refers to
XX	
CC	polynucleotide sequences that exhibit differential expression patterns in
XX	
CC	cells activated by the transactivator MYCN, where MYCN is a proto-
XX	
CC	oncogene that is amplified in neuroblastoma cells and is common in small
XX	
CC	cell lung cancers. The present invention describes these cDNA molecules
XX	
CC	as useful for in hybridisation assays to detect expression of nucleic
XX	
CC	acids (or complementary nucleic acids) in a present in a given sample, as
XX	
CC	well as for screening assays by identifying molecules or compounds that
XX	
CC	specifically bind the cDNA as a ligand and modulate function or activity.
XX	
CC	Accordingly, these compositions exhibit cytostatic activity and can also
XX	
CC	be used for gene therapy purposes. This polynucleotide sequence is a cDNA
XX	
CC	that is differentially expressed in MYCN activated cells, given in an
XX	
CC	exemplification of the invention. NOTE: This sequence does not appear in
XX	
CC	the printed specification but has been obtained in electronic format from
XX	
CC	the US Patent Office at
XX	
CC	ftp://seqdata.uspto.gov/sequence.html?DocID=20030119009.
XX	
SEQ	Sequence 3917 BP; 1232 A; 863 C; 843 G; 979 T; 0 U; 0 Other;
	Query Match 3.5%; Score 54.2; DB 10; Length 3917;
	Best Local Similarity 61.9%; Pred. No. 0.071;
	Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY	528 AATGAGACGAGTGTGCGGTGTGCGGAGCGCGGAGGACTCTGCTGACGGCTGC 587
DB	2359 AATGAGACTGTGTGTGCAAGTTTGTCAAAACGAGGGGAATCTCTCTGTGTGAAAGTGC 2418
QY	588 CTTGCGGCGCTTTCACCTGCGCCTGTCCCTTCGCTCGGAGAGATCCCAAGTGGAAC 647
DB	2419 CCCAAAGTATTCCTCTTCTTGTGCAATGTGCGCCACATGTGCAAAATTTTCCAAAGTGGAG 2478
QY	648 TGGAGTGCTCCAGCTGCC 666
DB	2479 TGGATTGTGCACTTTCTGCG 2497
RESULT 47	
AAK94864	
ID	AAK94864 standard; cDNA; 3997 BP.
XX	
AAK94864;	
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human full-length cDNA, SEQ ID NO: 4044.
XX	
KW	Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX	
OS	Homo sapiens
XX	

PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-0018774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
PI Oca T, Nishikawa T, Isogai T, Hayaishi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR F-PSDB; AAM93902.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
PS Claim 8; SEQ ID NO 4044; 1380bp + Sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5' and 3' ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesized by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from EPO
XX
SQ Sequence 3997 BP; 1221 A; 944 C; 911 G; 921 T; 0 U; 0 Other;
XX
Query Match 3.5%; Score 54.2; DB 4; Length 3997;
Best Local Similarity 61.9%; Pred. No. 0.071;
Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 528 AATGAGACGAGTGTGCGGACGCGGAGCTCATCTGCTGACGCGCTGC 587
DB 2688 AATGAGACGAGTGTGCGGACGCGGAGCTCATCTGCTGACGCGCTGC 2747
QY 588 CCTCGGCGCTTCCACTGCGCTGCTGCTCCCTCCGCTCGGAGATCCCACTGGAC 647
DB 2748 CCCAAAGTATTCATCTTCTGTGATGCGCCACATTGACAAATTTTCAAGTGAGAG 2807
QY 648 TGGAGGTGCTCCAGCTGCC 666
DB 2808 TGGATTGCACTTTCTGCC 2826
XX
RESULT 48
ADL32011
ID ADL32011 standard; cDNA; 3997 BP.
XX
AC ADL32011;
XX
DT 20-MAY-2004 (first entry)
XX
DE Full length human cDNA clone SegID 4044.
XX
KW human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method; ss; gene.
XX
OS Homo sapiens.
XX
PN EP1396543-A2.
XX
PD 10-MAR-2004.
XX

XX
PF 07-JUL-2000; 2003EP-00025638.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-0018774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
PI Oca T, Nishikawa T, Isogai T, Hayaishi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2004-204755/20.
DR F-PSDB; ADL32012.
XX
PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX
PS Example 1; SEQ ID NO 4044; 1340bp; English.
XX
CC This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3'
CC ends using the oligo-capping method. This polynucleotide sequence is a
XX full length human cDNA clone of the invention.
XX
SQ Sequence 3997 BP; 1221 A; 944 C; 911 G; 921 T; 0 U; 0 Other;
XX
Query Match 3.5%; Score 54.2; DB 12; Length 3997;
Best Local Similarity 61.9%; Pred. No. 0.071;
Matches 86; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 528 AATGAGACGAGTGTGCGGACGCGGAGCTCATCTGCTGACGCGCTGC 587
DB 2688 AATGAGACGAGTGTGCGGACGCGGAGCTCATCTGCTGACGCGCTGC 2747
QY 588 CCTCGGCGCTTCCACTGCGCTGCTGCTCCCTCCGCTCGGAGATCCCACTGGAC 647
DB 2748 CCCAAAGTATTCATCTTCTGTGATGCGCCACATTGACAAATTTTCAAGTGAGAG 2807
QY 648 TGGAGGTGCTCCAGCTGCC 666
DB 2808 TGGATTGCACTTTCTGCC 2826
XX
RESULT 49
AAK53491/C
ID AAK53491 standard; DNA; 114955 BP.
XX
AC AAK53491;
XX
DT 05-JUL-1999 (first entry)
XX
DE Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
KW Antisense oligonucleotide; multiple target; antisense treatment;
KW impaired respiration; inflammation; lung disease;
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW acute asthma; allergy; asthma; impeded respiration;
KW respiratory distress syndrome; pain; cystic fibrosis;
KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW prostate cancer; ss.
XX
OS Synthetic.
XX

PN WO913886-A1.
 XX 25-MAR-1999.
 XX 17-SEP-1998; 98WO-US019419.
 XX 17-SEP-1997; 97US-0059160P.
 PR 09-JUN-1998; 98US-00093972.
 XX (UYEC-) UNIV EAST CAROLINA.
 XX NYCE JW;
 DR WPI; 1999-229400/19.
 XX New antisense oligonucleotides used in treatment of, e.g. pulmonary
 PT vasoconstriction.
 XX
 PS Disclosure; Page 37; 120pp; English.
 CC The specification describes antisense oligonucleotides (AA52869-X55271)
 CC directed against at least 2 mRNAs selected from target genes, coding and
 CC non-coding regions of RNAs corresponding to target genes, gene initiation
 CC codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-
 CC end and the juxta-section between coding and non-coding regions and all
 CC segments of RNAs encoding proteins associated with one or more diseases,
 CC conditions or mixtures. The antisense oligonucleotides may be derived
 CC from sequences AA55272-74. These multiple target oligonucleotides
 CC (specifically AA55180-271) can be used for the antisense treatment of
 CC diseases and conditions. Typical diseases and conditions are those
 CC associated with impaired respiration and inflammation, including lung
 CC diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,
 CC acute asthma, allergies, asthma, impeded respiration, respiratory
 CC distress syndrome, pain, cystic fibrosis, pulmonary hypertension,
 CC pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary
 CC disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.
 CC colon cancer, breast cancer, lung cancer, pancreatic cancer,
 CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as
 CC well as all types of cancers which may metastasize or have metastasized
 CC to the lungs, including breast and prostate cancer
 XX
 SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other:
 Query Match 3.5%; Score 54; DB 2; Length 114955;
 Best Local Similarity 30.8%; Pred. No. 0.12;
 Matches 226; Conservative 63; Mismatches 443; Indels 2; Gaps 1;
 QY 548 GGTGGGAGCGGGGAGAGTATCTGTGACGGCTGCGGCGCTTCACACTGGC 607
 DB 105539 GGGCGGAGCGAGGGGCCNNHNNNSVCGAGCTVCGCGCCNNHNNNSCGGCCGCGC 105480
 QY 608 CTGCTGTCTCCCTCCGCTCCGAGATCCCAAGTGGAGCTGTCCAGTGGCT 667
 DB 105479 HNNNSCGGCGCGCGGCGNNHNNNSCGGCCGCGGCGGNNHNNNSCGGCCGCGC 105420
 QY 668 GCAAGCAAGTTCAGAGAGTGCAGCCCGGCGAGAGAGAGAGAGAGAGAGAGAGAG 727
 DB 105419 NSCGGCGCGCGCGCGGCGNNHNNNSCGGCGCGGCGGCGGCGGCGGCGGCGG 105360
 QY 728 CGTGGAGAGCGCGCTCGCGCGGCGGCTTGTAGTGGGAGAGAGAGAGAGAGAGAGAG 787
 DB 105359 CNNHNNNSCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105300
 QY 788 TGGGGAACCCCTAGCGGAGATGACAGACTTGTCTACAAGACACTCGGCTCGGC 847
 DB 105299 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105240
 QY 848 TTCTGAGCGCGCGCTCCAGAGTGTGACTTC--TCGCGCTTGAACCCCTACTGTGTG 905
 DB 105239 GCGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105180
 QY 906 GGCTCTGAGAGGTACAGAAAGTGTCTGTGTGCGCGGCTGTGCGGAGATGTG 965

DB 105179 GCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105120
 QY 966 ACGAGCTGTCTCGGTGTGTACTGACGCGCGGTCTTCACTGGCGGTGCACTTCCGA 1025
 DB 105119 CGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105060
 QY 1026 GCGCGGAGCTCCCGGCGGCGGAGAGCGGCGGTCTGCAATCTGTCTCAGAGACGTGACC 1085
 DB 105059 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105000
 QY 1086 CCAGCGCTGTGAGAGGAGGTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1145
 DB 104999 NSCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104940
 QY 1146 AAGATGACACTGACGAGTACGAGCGGCTGTGCAAGGAGATGACTGTGCTCTTG 1205
 DB 104939 VGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104880
 QY 1206 AGCGAGACACTTGTGATGTGATCTCTGAGTGGGCGGCGGCGGCGGCGGCGGCGG 1265
 DB 104879 GGNHNNNSVGGCGGCGGNNHNNNSVGGCGGCGGNNHNNNSVGGCGGCGGNNH 104820
 QY 1266 GCCCGCTTCCCTC 1279
 DB 104819 NNSCCVGGCGGCGG 104806
 RESULT 50
 AA125090/c
 ID AA125090 brandard; DNA; 292 BP.
 XX
 AC AA125090;
 XX
 DT 12-OCT-2001 (first entry)
 DE Probe #15023 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; 88.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000670.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human cervical epithelial cells.
 XX
 PS Claim 25; SEQ ID NO 15023; 487bp; English.
 CC
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging of

CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pcc_sequences

XX Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 292;
Best Local Similarity 61.8%; Pred. No. 0.098;

Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGTGTGTCGAGCGGCGGAGCTCATCTGTGACGCTGCC 588

DB 140 ATATGGAATTCTGTGGGCTGTGCAAGATGTGGGAACTGCTCTGTGATTAAGTTC 81

QY 589 CTCGGGCTTCCACTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGACCT 648

DB 80 CTTCTTCTTACCACTTCACCTGCTGAATCCCTCACTTCAGAGATCCCAACGCTGAAT 21

QY 649 GGAGGTGCTCCAGCTG 664

DB 20 GGCTCTGTCCCGTTG 5

RESULT 51

ABA70776/c

ID ABA70776 standard; DNA; 292 BP.

AC ABA70776;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #19081.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000669.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PT WPI; 2001-483447/52.

PS Claim 4; SEQ ID NO 19081; 639pp + Sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring

CC human gene expression in a sample derived from human foetal liver. The

CC single exon nucleic acid probes may be used for predicting, measuring and

CC displaying gene expression in samples derived from human foetal liver. The

CC present sequence is a single exon nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

Query Match 3.4%; Score 52.8; DB 4; Length 292;
Best Local Similarity 61.8%; Pred. No. 0.098;

Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGTGTGTCGAGCGGCGGAGCTCATCTGTGACGCTGCC 588

DB 140 ATATGGAATTCTGTGGGCTGTGCAAGATGTGGGAACTGCTCTGTGATTAAGTTC 81

QY 589 CTCGGGCTTCCACTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGACCT 648

DB 80 CTTCTTCTTACCACTTCACCTGCTGAATCCCTCACTTCAGAGATCCCAACGCTGAAT 21

QY 649 GGAGGTGCTCCAGCTG 664

DB 20 GGCTCTGTCCCGTTG 5

RESULT 52

AA150949/c

ID AA150949 standard; DNA; 292 BP.

AC AA150949;

DT 17-OCT-2001 (first entry)

DE Probe #19635 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;

OS genetic disorder; ss.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PT WPI; 2001-488897/53.

PS Claim 25; SEQ ID NO 19635; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders

QY 529 ATGAGACAGTGTGCGGTGTGTCGAGCGGCGGAGCTCATCTGTGACGCTGCC 588

DB 140 ATATGGAATTCTGTGGGCTGTGCAAGATGTGGGAACTGCTCTGTGATTAAGTTC 81


```
RESULT 55
AAK19029/c
ID AAK19029 standard; DNA; 292 BP.
XX
AC AAK19029;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 19020.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX
PS Example 4; SEQ ID NO 19020; 650bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 292;
Best Local Similarity 61.8%; Pred. No. 0.098;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGAGCAGGTGCGGTGTCGGAGCGGGAGGCTCATCTGCTGTGACGGCTGCC 588
DB 140 ATATGGAATTCTGTGCGGTCTGCAAGAGTGTGGGAACTGCTGTGTGATTACTGTTC 81
QY 589 CTCGGGCTTTCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGCT 648
DB 80 CTTCTTCTTAACCAATCACTGCTGAATCCCACTTCAGAGATCCCAAGGTGAAT 21
QY 649 GGAAGTGTCTCAGCTG 664
DB 20 GGCTCTGTCCCGCTTG 5

RESULT 56
ABS44646/c
ID ABS44646 standard; DNA; 292 BP.
```

```
XX
AC ABS44646;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID NO 19636.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 4; SEQ ID NO 19636; 658bp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (i) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (i) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 292;
Best Local Similarity 61.8%; Pred. No. 0.098;
Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGAGCAGGTGCGGTGTCGGAGCGGGAGGCTCATCTGCTGTGACGGCTGCC 588
DB 140 ATATGGAATTCTGTGCGGTCTGCAAGAGTGTGGGAACTGCTGTGTGATTACTGTTC 81
QY 589 CTCGGGCTTTCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGCT 648
DB 80 CTTCTTCTTAACCAATCACTGCTGAATCCCACTTCAGAGATCCCAAGGTGAAT 21
QY 649 GGAAGTGTCTCAGCTG 664
DB 20 GGCTCTGTCCCGCTTG 5

RESULT 57
```

ABSI9225/c
 ID ABSI9225 standard; DNA; 292 BP.
 XX
 AC ABSI9225;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon probe ORF from lung SEQ ID NO 19216.
 XX
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary hemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000665.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 XX
 PR 26-MAY-2000; 2000US-0207456P.
 XX
 PR 30-JUN-2000; 2000US-00608408.
 XX
 PR 03-AUG-2000; 2000US-00632366.
 XX
 PR 21-SEP-2000; 2000US-0234687P.
 XX
 PR 27-SEP-2000; 2000US-0236359P.
 XX
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples.
 XX
 PS Claim 4; SEQ ID NO 19216; 634bp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC; the novel set of probes which hybridize at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a sample
 CC collected from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of the
 CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarray having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene expression
 CC analysis, and for identifying exons in a gene, particularly using human
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung

CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
 CC Karsagen syndrome, fibrocystic pulmonary dysplasia, primary ciliary
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
 CC present sequence is a single exon probe open reading frame of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SO Sequence 292 BP; 65 A; 94 C; 67 G; 66 T; 0 U; 0 Other;
 XX
 Query Match 3.4%; Score 52.8; DB 6; Length 292;
 Best Local Similarity 61.8%; Pred. No. 0.098;
 Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 XX
 QY 529 ATGAGACGAGTGTCCGCTGTCGAGACGCGGAGCTCATCTGCTGACGCTGCC 588
 DB 140 ATATGGAAATTCGTCCGCTGTCAAGAGATGTGGGAACTGCTGTGTGATCCTGTC 81
 QY 589 CTCGGGCTTCCACCTGCGCTGCTCCCTCCGCTCCGAGATCCCACTGGAGCCT 648
 DB 80 CTTCTTCTACCACTCACTGCTGAATCCCACTTCCAGATCCCAACGGTGAAT 21
 QY 649 GGAGGTCTCCAGCTG 664
 DB 20 GGCTGTGCTCCGCTTG 5
 XX
 RESULT 58
 AAI15873/c
 ID AAI15873 standard; DNA; 476 BP.
 XX
 AC AAI15873;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Probe #5806 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 XX
 OS Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000670.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 XX
 PR 26-MAY-2000; 2000US-0207456P.
 XX
 PR 30-JUN-2000; 2000US-00608408.
 XX
 PR 03-AUG-2000; 2000US-00632366.
 XX
 PR 21-SEP-2000; 2000US-0234687P.
 XX
 PR 27-SEP-2000; 2000US-0236359P.
 XX
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human cervical epithelial cells.
 XX
 PS Claim 25; SEQ ID NO 5806; 487bp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SEN). The present sequence is one such probe. The SENPs are derived

CC from human HeLa cells. The SENSEs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;

XX Query Match 3.4%; Score 52.8; DB 4; Length 476;
XX Best Local Similarity 61.8%; Pred. No. 0.1;
XX Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGTGTCGGACGCGGGAGCTCATCTGCTGACGCTGCC 588
DB 428 ATATGGAATTCTGTGCGGTCTGCAAGATGTGGAGACTGCTGTGTGATACCTGTC 369
QY 589 CTCGGGCGCTTCCACCTGGCTGCTGCCCTCCCTCCGTCGGAGATCCCACTGGAGACT 648
DB 368 CTTCTTCTTACCACTTCACCTGCTGATTCCTCCACTTCAGAGATCCCAAGCTGTAAT 309
QY 649 GGAGGTGCTCCAGCTG 664
DB 308 GGCTCTGTCCCGTTG 293

RESULT 59
ABAS8175/c
ID ABAS8175 standard; DNA; 476 BP.
XX
XX ABA58175;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #6480.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX MO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human foetal liver.
XX
XX Claim 1; SEQ ID NO 6480; 639bp + Sequence Listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human foetal liver. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;

XX Query Match 3.4%; Score 52.8; DB 4; Length 476;
XX Best Local Similarity 61.8%; Pred. No. 0.1;
XX Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGTGTCGGACGCGGGAGCTCATCTGCTGACGCTGCC 588
DB 428 ATATGGAATTCTGTGCGGTCTGCAAGATGTGGAGACTGCTGTGTGATACCTGTC 369
QY 589 CTCGGGCGCTTCCACCTGGCTGCTGCCCTCCGTCGGAGATCCCACTGGAGACT 648
DB 368 CTTCTTCTTACCACTTCACCTGCTGATTCCTCCACTTCAGAGATCCCAAGCTGTAAT 309
QY 649 GGAGGTGCTCCAGCTG 664
DB 308 GGCTCTGTCCCGTTG 293

RESULT 60
AA137782/c
ID AA137782 standard; DNA; 476 BP.
XX
XX AA137782;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #6468 used to measure gene expression in human placenta sample.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX Claim 25; SEQ ID NO 6468; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders

XX SQ Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;

XX Query Match 3.4%; Score 52.8; DB 4; Length 476;
XX Best Local Similarity 61.8%; Pred. No. 0.1;
XX Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACGAGTGTGCGGTCGCGGACGCGGAGACTCATCTGCTGTGACGCTGCC 588
 DB 428 ATATGGAATTCTGCGGTCGCGGACGAGATGTGGGAACTGCTGTGATTAACCTGTC 369
 QY 589 CTCGGGCTTTCCACCTGAGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGCT 648
 DB 368 CTTCTTCTTCCACCATTCATCCTGTAATCCCACTTCCAGAGATCCCAAGTGAAT 309
 QY 649 GGAGGTGCTCCAGCTG 664
 DB 308 GGCTGTGCTCCGCTTG 293

RESULT 61
 ABA27360/c
 ID ABA27360 standard; DNA; 476 BP.

AC ABA27360;
 XX
 XX 23-JAN-2002 (first entry)
 DT
 XX
 XX Probe #5826 for gene expression analysis in human heart cell sample.
 DE
 XX Human; gene expression; heart; microarray; vascular system; probe;
 KM cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200157274-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US000666.
 XX
 XX 04-FEB-2000; 2000US-0180312P.
 XX
 XX 26-MAY-2000; 2000US-0207456P.
 XX
 XX 30-JUN-2000; 2000US-00608408.
 XX
 XX 03-AUG-2000; 2000US-00632366.
 XX
 XX 21-SEP-2000; 2000US-0234687P.
 XX
 XX 27-SEP-2000; 2000US-0236359P.
 XX
 XX 04-OCT-2000; 2000GB-00024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-488990/53.
 XX
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts.
 XX
 XX Claim 1; SEQ ID NO 5826; 530bp; English.
 XX
 XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease. Note: The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;
 SO

Query Match 3.4%; Score 52.8; DB 4; Length 476;
 Best Local Similarity 61.8%; Pred. No. 0.1;
 Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACGAGTGTGCGGTCGCGGACGCGGAGACTCATCTGCTGTGACGCTGCC 588
 DB 428 ATATGGAATTCTGCGGTCGCGGACGAGATGTGGGAACTGCTGTGATTAACCTGTC 369
 QY 589 CTCGGGCTTTCCACCTGAGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGCT 648
 DB 368 CTTCTTCTTCCACCATTCATCCTGTAATCCCACTTCCAGAGATCCCAAGTGAAT 309
 QY 649 GGAGGTGCTCCAGCTG 664
 DB 308 GGCTGTGCTCCGCTTG 293

RESULT 62
 AAK31915/c
 ID AAK31915 standard; DNA; 476 BP.

AC AAK31915;
 XX
 XX 06-NOV-2001 (first entry)
 DT
 XX
 XX Human bone marrow expressed single exon probe SEQ ID NO: 6472.
 DE
 XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KM microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200157276-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US000666.
 XX
 XX 04-FEB-2000; 2000US-0180312P.
 XX
 XX 26-MAY-2000; 2000US-0207456P.
 XX
 XX 30-JUN-2000; 2000US-00608408.
 XX
 XX 03-AUG-2000; 2000US-00632366.
 XX
 XX 21-SEP-2000; 2000US-0234687P.
 XX
 XX 27-SEP-2000; 2000US-0236359P.
 XX
 XX 04-OCT-2000; 2000GB-00024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-488990/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human bone marrow.
 XX
 XX Example 4; SEQ ID NO 6472; 658bp + Sequence Listing; English.
 XX
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention
 XX
 XX Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;
 SO

Query Match 3.4%; Score 52.8; DB 4; Length 476;
 Best Local Similarity 61.8%; Pred. No. 0.1;
 Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACGAGTGTGCGGTCGCGGACGCGGAGACTCATCTGCTGTGACGCTGCC 588
 DB 428 ATATGGAATTCTGCGGTCGCGGACGAGATGTGGGAACTGCTGTGATTAACCTGTC 369
 QY 589 CTCGGGCTTTCCACCTGAGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGCT 648
 DB 368 CTTCTTCTTCCACCATTCATCCTGTAATCCCACTTCCAGAGATCCCAAGTGAAT 309

QY 649 GGAGGCTCCAGCTG 664
 DB 308 GGCTCTGCCGCTTG 293

RESULT 63
 ID AAK06255/c
 AAK06255 standard; DNA; 476 BP.

AAK06255;
 05-NOV-2001 (first entry)

Human brain expressed single exon probe SEQ ID NO: 6246.

Human; brain expressed exon; gene expression analysis; probe; microarray;
 Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
 ss.

Homo sapiens.

WO200157275-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000667.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-483446/52.

Single exon nucleic acid probes for analyzing gene expression in human brains.

Example 4; SEQ ID NO 6246; 650bp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention

Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 476;

Best Local Similarity 61.8%; Pred. No. 0.1;

Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGAGCGGAGGAGCTCATCTGTGACGGCTGCC 588

DB 428 ATATGAATTTCTGTGCGGCTGTCAAGATGTGGGAATGCTCTGTGATACCTGTC 369

QY 589 CTCGGGCTTTCACCTGCGCTGTCCCTTCGCTCCGGAGATCCCAAGTGGACCT 648

DB 368 CTTCTTCTTACCACTGCTGCTGAATCCCACTTCAGAGATCCCAAGTGAAT 309

QY 649 GGAGGCTCCAGCTG 664

DB 308 GGCTCTGCCGCTTG 293

RESULT 64
 ID ABS31605/c
 ABS31605 standard; DNA; 476 BP.

ABS31605;

25-FEB-2003 (first entry)

Human liver single exon probe, SEQ ID No 6595.

Human; single exon nucleic acid probe; liver; cirrhosis;
 hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
 coronary heart disease; ss.

Homo sapiens.

WO200157273-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000664.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488898/53.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human adult liver.

Claim 1; SEQ ID NO 6595; 658bp; English.

The invention relates to a single exon nucleic acid probe (SENP) (1) for measuring human gene expression in a sample derived from human adult liver, comprising one of 1109 defined nucleotide sequences given in the specification (or complements/ fragments). The probe hybridises at high stringency to a nucleic acid molecule expressed in the human adult liver. (1) may be used for predicting, measuring and displaying gene expression in samples derived from human adult liver. The genes identified may be involved in genetic liver diseases such as cirrhosis, hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is associated with coronary heart disease. ABS25011-ABS51005 represent human liver single exon nucleic acid probes of the invention. Note: The sequence information for this patent does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 476 BP; 131 A; 132 C; 103 G; 110 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 476;

Best Local Similarity 61.8%; Pred. No. 0.1;

Matches 84; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 529 ATGAGACAGTGTGCGGAGCGGAGGAGCTCATCTGTGACGGCTGCC 588

DB 428 ATATGAATTTCTGTGCGGCTGTCAAGATGTGGGAATGCTCTGTGATACCTGTC 369

QY 589 CTCGGGCTTTCACCTGCGCTGTCCCTTCGCTCCGGAGATCCCAAGTGGACCT 648

DB 368 CTTCTTCTTACCACTGCTGCTGAATCCCACTTCAGAGATCCCAAGTGAAT 309

QY 649 GGAGGCTCCAGCTG 664

DB 308 GGCTCTGCCGCTTG 293

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-
CC ABB82072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 6025 BP; 1591 A; 1529 C; 1756 G; 1149 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 6025;
Best Local Similarity 54.7%; Pred. No. 0.15;

Matches 105; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 522 CAGAAATAGAGACGAGTGTGCTGTCGGAGCGCGGAGCTCATCTGCTGTAC 581
DB 1093 CATGAGCACACGAGACTACTGCGAGGTGTCCAGCAGAAGAGTGAGATCCTGTGCGAC 1152
QY 582 GGCTGCGCTCGGGCCCTTCCACCTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCGCAGT 641
DB 1153 ACCTGCGCTCGGGCATATCACTTGTGTCTGAGCCAGAACTCGATTAACCGCCAGAG 1212
QY 642 GGGACCTGAGGTGTCTCCAGCTGCTGCGAGCAAGTCAGAGGTGCGACCCCGGCGCA 701
DB 1213 GCGAAGTGTGCTGTGCTCGCACTGCGAGGCTGACGAGGTGCTGTGAGGAAGAGACGAT 1272
QY 702 GAGAGACCCCGG 713
DB 1273 GATGAGCACACAG 1284

RESULT 67

AB129648/c
ID ABL29648 standard; DNA; 7303 BP.

AC ABL29648;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40417.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; de.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.

PS Claim 1; SEQ ID NO 40417; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-
CC ABB82072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 7303 BP; 1509 A; 2002 C; 1786 G; 2006 T; 0 U; 0 Other;

Query Match 3.4%; Score 52.8; DB 4; Length 7303;
Best Local Similarity 54.7%; Pred. No. 0.15;

Matches 105; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 522 CAGAAATAGAGACGAGTGTGCTGTCGGAGCGCGGAGCTCATCTGCTGTAC 581
DB 5509 CATGAGCACACGAGACTACTGCGAGGTGTCCAGCAGAAGAGTGAGATCCTGTGCGAC 5450
QY 582 GGCTGCGCTCGGGCCCTTCCACCTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCGCAGT 641
DB 5449 ACCTGCGCTCGGGCATATCACTTGTGTGCTGAGCCAGAACTCGATTAACCGCCAGAG 5390
QY 642 GGGACCTGAGGTGTCTCCAGCTGCTGCGAGCAAGTCAGAGGTGCGACCCCGGCGCA 701
DB 5389 GCGAAGTGTGCTGTGCTCGCACTGCGAGGCTGACGAGGTGCTGTGAGGAAGAGACGAT 5330
QY 702 GAGAGACCCCGG 713
DB 5329 GATGAGCACACAG 5318

RESULT 68

AB099469
ID AB099469 standard; cDNA; 4653 BP.

AC AB099469;

XX 25-FEB-2003 (first entry)

DE Human coding sequence SEQ ID 202.

XX Human; expressed sequence tag; EST; chromosome 1;
KW haematopoietic disorder; central nervous system disease; viral infection;
KW peripheral nervous system disease; non-healing wound; infectious disease;
KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW fungal infection; autoimmune disorder; coagulation disorder; noctropic;
KW anti-allergic; antineoplastic; immunosuppressive; neuroprotective;
KW cytosolic; haemostatic; virucide; antibacterial; fungicide;
KW immunostimulant; cerebroprotective; gene therapy; gene; ss.

OS Homo sapiens.

XX WO200259260-A2.

PD 01-AUG-2002.

XX 16-NOV-2001; 2001WO-US042950.

PR 17-NOV-2000; 2000US-00714936.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;

DR WPI; 2002-590824/63.

PT New isolated polynucleotide, useful in research, diagnostic or
XX therapeutic methods, e.g. preventing or treating disorders involving
XX aberrant protein expression or biological activity.

PS Claim 1; SEQ ID NO 202; 394bp; English.

CC The present invention relates to novel human coding sequences (AB099268-
XX

CC AB099608) and proteins (ABP64682-ABP65022). The sequences are useful in
CC therapeutic, diagnostic and research methods. The polynucleotides may be
CC used in the field of molecular biology as hybridisation probes, primers
CC for PCR, for chromosome and gene mapping, for the recombinant production
CC of protein, or in generation of anti-sense DNA or RNA. The
CC polynucleotides are useful in diagnostics as expressed genes tags
CC (ESTs) for identifying expressed genes or for physical mapping of the
CC human genome. The proteins may be used as molecular weight markers, or as
CC nutritional sources or supplements. The proteins may be used to maintain
CC and expand cell population in a totipotent or pluripotent state
CC useful for re-engineering damaged or diseased tissues, transplantation,
CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The
CC polynucleotides and proteins are useful for preventing, treating or
CC ameliorating disorders involving aberrant protein expression or
CC biological activity, e.g. haematopoietic disorders, central/peripheral
CC nervous system diseases, mechanical and traumatic disorders, non-healing
CC wounds, immune deficiencies and disorders, infectious diseases caused by
CC viral, bacterial or fungal infection, autoimmune disorders, allergic
CC reactions and conditions, coagulation disorders, or cancer. The
CC polynucleotide sequences of the invention were assembled from ESTs
CC isolated mainly by sequencing by hybridisation, and in some cases,
CC sequences obtained from one or more public databases. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 4653 BP; 1097 A; 1368 C; 1416 G; 772 T; 0 U; 0 Other;

XX Query Match 3.4%; Score 52.6; DB 6; Length 4653;

XX Best Local Similarity 66.1%; Pred. No. 0.16;

XX Matches 76; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 533 GGACGAGTGTGCGGTGTGCGGAGCGGAGCTCATCTGTGTGACGGCTCCGCG 592
DB 8 GGAGTCTCTCCGCGGTGTGCGGAGCGGAGCTCATCTGTGTGACGGCTCCGCG 67

OY 593 GGCTTCGACGCTGCGCTGCTGCGGAGGATCCGAGTGGGAC 647
DB 68 CTCCTACGACCTGATTCGCTCAACCGCGCTGCGGAGGATCCGAGTGGGAC 122

RESULT 69

ABV44220 standard; cDNA; 539 BP.

XX ABV44220;

XX 16-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 44211.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.

XX 17-FEB-2000; 2000US-0183319P.

XX 16-MAR-2000; 2000US-0189862P.

XX 25-MAY-2000; 2000US-0207454P.

XX 09-JUN-2000; 2000US-0211314P.

XX 18-JUL-2000; 2000US-0219007P.

XX 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JB;

XX

DR WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

XX prostate cells and correlating with presence of prostate cancer, useful

XX for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 8783; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (1) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (1) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 539 BP; 125 A; 148 C; 150 G; 115 T; 0 U; 1 Other;

XX Query Match 3.4%; Score 52.2; DB 5; Length 539;

XX Best Local Similarity 66.4%; Pred. No. 0.14;

XX Matches 75; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

OY 553 GGACGCGCGGAGCTCATCTGTGACGCGCTGCGGAGCTTCACCTGCGCTGCC 612
DB 50 GGAGCGCGCGGAGCTCATCTGTGACGCGGAGCTTCACCTGCGCTTCATTCATTC 109

OY 613 TGTCCCTCTCCGCGGAGATCCGAGTGGGAGCTGAGGCTTCACCTGC 665
DB 110 TTAACCTCTCCGCTGCTGACATTCGCAATGGAATGCTGTGCTCCCGATGC 162

RESULT 70

ABV29094 standard; cDNA; 539 BP.

XX ABV29094;

XX 16-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 29085.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.

XX 17-FEB-2000; 2000US-0183319P.

XX 16-MAR-2000; 2000US-0189862P.

XX 25-MAY-2000; 2000US-0207454P.

XX 09-JUN-2000; 2000US-0211314P.

XX 18-JUL-2000; 2000US-0219007P.

XX 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JB;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

XX prostate cells and correlating with presence of prostate cancer, useful

XX for detecting presence of prostate cancer, stage of prostate cancer.

XX	Claim 1; Page 6175; 11750pp; English.
PS	
XX	
CC	The invention relates to an isolated nucleic acid molecule (1) comprising
CC	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC	specification or its complement. (1) is useful for: (a) assessing whether
CC	a patient is afflicted with prostate cancer; (b) monitoring the
CC	progression of prostate cancer in a patient; (c) assessing the efficacy
CC	of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC	the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;
CC	(f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC	determining whether prostate cancer has metastasized in a patient; (h)
CC	assessing the aggressiveness or indolence of prostate cancer in a patient
CC	; (1) is also useful as a pharmacodynamic or pharmacogenomic marker
XX	
XX	
SQ	Sequence 539 BP; 125 A; 148 C; 147 G; 113 T; 0 U; 6 Other;
QY	Query Match 3.4%; Score 52.2; DB 5; Length 539;
Db	Best Local Similarity 66.4%; Pred. No. 0.14; Mismatches 38; Indels 0; Gaps 0
QY	Matches 75; Conservative 0;
Db	553 GGGAGGCGGGGAGCTCATCTGCTGTGAGCGCTGCCCTGGGCTTCACCTGAGCTGCC 612
	50 GGGAGGCGGGGAGCTCATCTGCTGTGAGCGCGCTGCATCTCTTCACCATTCATTCATGTC 109
	613 TGTCCCTCTCGCTCCGGAGATCCCGAGTGGGACCTGAGAGTGCCTCACCTGC 665
	110 TAAACCTCTCCCTGCTGACATTCCTCCAAATGAGTGAATGGCTGTGTCTCCCATGC 162
RESULT 71	
ADCC30621	
ID	ADCC30621 standard; cDNA; 1670 BP.
XX	
XX	
AC	ADCC30621;
DT	
XX	18-DEC-2003 (first entry)
DE	
XX	Human novel cDNA sequence, SEQ ID NO:703.
XX	
KW	Human; diagnostic; drug screening; forensics; gene mapping;
KW	biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW	neurodegenerative diseases; anemia; platelet disorder; wound; burns;
KW	ulcers; osteoporosis; autoimmune disease; cancer;
KW	molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW	neuroprotective; anemic; antianemic; anticoagulant; thrombolytic; vulnerary;
KW	antitumor; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW	gene therapy; chromosome 22; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO2003029271-A2.
XX	
PD	10-APR-2003.
XX	
PF	24-SEP-2002; 2002WO-US030474.
XX	
PR	24-SEP-2001; 2001US-0324631P.
XX	
PA	(HYSE-) HYSEQ INC.
PI	
PI	Tang TV, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI	Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI	Haley-Vicente D, Drmanac RT;
XX	
DR	WPI; 2003-371981/35.
XX	
DR	P-PSDB; ADC31592.
XX	
PT	New polynucleotide and polypeptide useful for diagnosing, preventing or
PT	treating conditions such as neurodegenerative diseases, anemias, platelet
PT	disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT	cancer.

XX		Claim 1; SEQ ID NO 703; 1185bp; English.
XX		
PX		The invention relates to 971 novel human cDNA sequences (ADC29919-
XX		ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC		invention also relates to nucleic acid sequences over 99% identical with
CC		the novel human cDNAs. The invention additionally encompasses expression
CC		vectors and host cells comprising a nucleic acid of the invention; the
CC		recombinant production of a polypeptide of the invention; an antibody
CC		against a polypeptide of the invention; a method of detecting
CC		polynucleotides or polypeptides of the invention; and methods of
CC		identifying a compound which binds to a polypeptide of the invention. The
CC		invention further discloses methods of preventing, treating or
CC		ameliorating a medical condition; kits comprising polynucleotide probes
CC		and/or monoclonal antibodies for carrying out the methods of the
CC		invention; methods for the identification of compounds that modulate the
CC		expression or activity of the polynucleotide and/or polypeptide; and 767
CC		coding sequences corresponding to the cDNA sequences of the invention
CC		(ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC		-ADC33394). The nucleic acids and polypeptides of the invention are
CC		useful in diagnostics, drug screening, forensics, gene mapping, in the
CC		identification of mutations responsible for genetic disorders or other
CC		traits, for assessing biodiversity, and in producing many other types of
CC		data and products dependent on DNA and amino acid sequences. They are
CC		also used for treating diseases such as Parkinson's disease, Alzheimer's
CC		disease and other neurodegenerative diseases, anaemia, platelet
CC		disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC		cancer. The nucleic acids may also be used as hybridisation probes or
CC		primers, and in the recombinant production of a protein. The polypeptides
CC		are also useful in generating antibodies, as molecular weight markers,
CC		and as food supplements. The present sequence represents a specifically
CC		claimed human cDNA sequence of the invention. Note: The sequence data for
CC		this patent did not form part of the printed specification, but was
CC		obtained in electronic format directly from WIPO at
CC		ftp.wipo.int/pub/published_pct_sequences.
XX		
XX		Sequence 1670 BP, 374 A, 596 C; 488 G; 212 T; 0 U; 0 Other;
Query Match	3.4%; Score 52; DB 10; Length 1670;	
Best Local Similarity	57.3%; Pred. No. 0.18;	
Matches	94; Conservative 0; Mismatches 70; Indels 0; Gaps 0;	
QY	529 ATGAGGACGATGTGGTCCCTGTGTTCGGACAGCGCGGGACTCATCTGCTGAACGGCTGCC	588
DB	1083 ACGATGACACATTGTGCCTGCGCAACGAGGAGGCCAACCTTGACGCTTCGGCACCTTGC	1144
QY	589 CTCGGGACCTTCCACCTGTGGCTGCTCCTCCCTCCGCTCCGGGAGATTCCTCCAGTAGACT	648
DB	1143 CGGGGGCCTTACCACCTCACTGCTGCTCGAGCCGCCCTTCAAGAAGCGCGCCAAAGCGCTGT	1200
QY	649 GGAGGTGCTCCAAGCTGCTGCGAAGGCAACAGTCCAGGAGGTGACG	692
DB	1203 GGGTGTGCCCCAGGTGCCAGACAAGAGGCTTTAAAGAAAAGACGAG	1246
RESULT 72		
AAK99602		
ID	AAK99602 standard; DNA, 1965 BP.	
XX		
AC	AAK99602;	
XX		
DT	10-APR-2003 (first entry)	
XX		
DE	MDMT related human DNA SEQ ID NO 66.	
XX		
XX	Cytostatic; antiatherosclerotic; osteopathic; antiarteriosclerotic;	
KW	hepatocytic; antiproliferatic; antiallergic; antianemic; antiaslathmic;	
KW	antihyproid; antiflammatory; antihelmintic; antidiabetic; nephrotropic;	
KW	ophthalmological; immunosuppressive; dermatological; antifungal;	
KW	antirheumatic; antizarthrnic; antibacterial; vrinucide; fungicide;	
KW	antiparasitic; protozoacide; tranquilliser; vulnerary; anti-HIV;	
KW	nootropic; neuroprotective; anticoagulant; cerebroprotective;	
KW	neuroleptic; molecules for disease detection and treatment; MDMT;	

XX immunogen; cancer; actinic keratosis; arteriosclerosis; atherosclerosis;
KM bursitis; cirrhosis; hepatitis; psoriasis; AIDS; rheumatoid arthritis;
KM adult respiratory distress syndrome; Addison's disease; allergy; anaemia;
KM asthma; osteoporosis; autoimmunity; haemolytic anaemia; scleroderma;
KM autoimmune thyroiditis; Crohn's disease; atopic dermatitis;
KM diabetes mellitus; Graves' disease; glomerulonephritis;
KM systemic lupus erythematosus; systemic sclerosis; ulcerative colitis;
KM haemodialysis; uveitis; trauma; Alzheimer's; Pick disease;
KM Parkinson disease; amyotrophic lateral sclerosis; epilepsy; stroke;
KM Huntington's disease; multiple sclerosis; dementia;
KM extrapyramidal disorder; motor neuron disorder; central nervous system;
KM neuromuscular disorder; metabolic; endocrine; toxic myopathy;
KM periodic paralysis; mental disorder; human; gene; de.
XX
OS Homo sapiens.
XX
XX WO200296951-A1.
XX
XX 05-DEC-2002.
XX
XX 24-MAY-2002; 2002WO-US016676.
XX
XX 25-MAY-2001; 2001US-0293723P.
PR 01-JUN-2001; 2001US-0295257P.
PR 08-JUN-2001; 2001US-0297220P.
PR 21-JUN-2001; 2001US-0300526P.
PR 29-JUN-2001; 2001US-0301874P.
PR 22-FEB-2002; 2002US-0359413P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
XX Tang TY, Yue H, Baughn MR, Duggan BM, Warren BA, Bandhan O,
PI Richardson TW, Burford N, Sanjanwala B, Becha SD, Yao MG, Yang J;
PI Tiran UK, Hafela AJA, Griffin JA, Swarnakar A, Elliott VS;
PI Reisporn SA, Khan FA, Lee EA, Yue H, Lu DM, Malia NK, *Thangavelu K;
PI Arizizu CS, Xu Y, Ison CH, Huang J, Ding L, Honchell CD,
PI Boroway ML, Emerling BK, Peterson DP, Lu Y, Ramkumar J, Mason PM,
PI Zeharajdan Y, Azimzal Y, Stuve LL, Kamigaki LL, Barroso I, Lee S;
PI Kable AE;
XX
XX WPI, 2003-140448/13.
XX P-PSDB; AAO26249.
XX
XX Novel molecules for disease detection and treatment and polynucleotide
PT encoding them useful for diagnosing, preventing or treating cell
PT proliferative, autoimmune/inflammatory, neurological and developmental
PT disorders.
XX
XX Claim 121; Page 248; 260pp; English.
PS
XX The invention relates to an isolated polypeptide chosen from molecules
CC for disease detection and treatment (MDPT) comprising a one of 39 114-
CC 1250 residue amino acid sequences, given in the specification, or a
CC biologically active or immunogenic fragment of the isolated polypeptide.
CC The isolated polypeptide is useful for screening a compound for
CC effectiveness as an agonist or antagonist of the isolated polypeptide.
CC The isolated polypeptide is also useful as an immunogen for preparing
CC polyclonal or monoclonal antibodies by hybridoma technology. The isolated
CC polypeptide and its encoding polynucleotide are useful for diagnosis,
CC treatment and prevention of cancer; actinic keratosis; arteriosclerosis,
CC atherosclerosis; bursitis; cirrhosis; hepatitis; psoriasis; AIDS; adult
CC respiratory distress syndrome; Addison's disease; allergic anaemia;
CC asthma; atherosclerosis; osteoporosis; autoimmune haemolytic anaemia,
CC autoimmune thyroiditis; Crohn's disease; atopic dermatitis; diabetes
CC mellitus; Graves' disease; glomerulonephritis; rheumatoid arthritis;
CC scleroderma; systemic lupus erythematosus; systemic sclerosis; ulcerative
CC colitis; haemodialysis; uveitis; viral, bacterial, fungal, parasitic,
CC protozoal, helminthic infections; trauma; Alzheimer's and Pick disease,
CC Parkinson disease, amyotrophic lateral sclerosis; epilepsy; stroke,
CC Huntington's disease, multiple sclerosis, dementia, and other
CC extrapyramidal disorder, motor neuron disorder, and other developmental
CC disorders of the central nervous system, neuromuscular disorders,
CC metabolic, endocrine and toxic myopathies, periodic paralysis, mental

CC		disorders including mood, anxiety and echinoprenic disorders, anaemia,
CC		renal tubular acidosis, epilepsy, hypothyroidism, glaucoma, sensorineural
CC		hearing loss and cataract. This polynucleotide sequence represents the
CC		DNA encoding a human MMDT protein relating to the invention
XX		
SQ	Sequence 1965 BP; 439 A; 664 C; 557 G; 305 T; 0 U; 0 Other;	
	Query Match	3.4%; Score 52; DB 10; Length 1965;
	Best Local Similarity	57.3%; Pred. No. 0.19;
	Matches	94; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
OY		
DB	529 ATGAGGACGAATGGCGGTGTGTGGGACGGGCGGAGAGCATCTGCCTGTGACGGCTGCC	588
	1085 ACGATGACACTGTGGCCGCTTGCAAGCAGAGGGGCGAACCTTCACACCCTGCGGACCTGCC	1144
OY	589 CTCGGGCGCTTCCACCTGGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCCAGTGGGACT	648
DB	1145 CGGGGGGCGCTTACCACTCAGCTGCGCTGGAGCGCGCCCTTAAGAAGCGCGCCCAAGGGCGGT	1204
OY	649 GGAAGTGCTCCAGCTGCCTCGCAGGCAACAGTCCAGAGAGTGACG	692
DB	1205 GGGTGTGCCCCAGGTGCCAGCAGAAAGGCTTTAAGAAAGACGAG	1248
RESULT 73		
ID	ADM02221	
XX	ADM02221 standard; CDNA; 2651 BP.	
XX	AC	
XX	ADM02221;	
DT	20-MAY-2004 (first entry)	
DE	Human CDNA of the invention SEQ ID NO:906.	
KW	ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.	
OS	Homo sapiens.	
PN	EP1347046-A1.	
PD	24-SEP-2003.	
PF	12-APR-2002; 2002EP-00008400.	
PR	22-MAR-2002; 2002JP-00137785.	
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.	
PI	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,	
PI	Yanamoto J, Isono Y, Hiro Y, Otsuka K, Nagai K, Irle R, Tamechika I,	
PI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maeno Y,	
DR	WPI; 2003-723558/69.	
XX	P-PADB; ADM04664.	
PT	New polynucleotides and polypeptides are useful in gene therapy, for	
PT	developing a diagnostic marker or medicines for regulating their	
PT	expression and activity, or as a target of gene therapy.	
ES	Claim 1; SEQ ID NO 906; 305pp; English.	
XX		
CC	The invention relates to a novel human polynucleotide and the encoded	
CC	polypeptide. A polynucleotide of the invention may have a use in gene	
CC	therapy. An oligonucleotide of the invention ADM062202-ADM06773 is useful	
CC	as a primer for synthesizing the polynucleotide or as a probe for	
CC	detecting the polynucleotide. The polynucleotides ADM0316-ADM03758 are	
CC	useful in gene therapy, for developing a diagnostic marker or medicines	
CC	for regulating their expression and activity, or as a target of gene	
CC	therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides	
CC	are useful as pharmaceutical agents. The present sequence represents a	
CC	cDNA sequence of the invention.	
SQ	Sequence 2651BP; 577 A; 730 C; 779 G; 565 T; 0 U; 0 Other;	

Query Match	3.4%	Score 52;	DB 11;	Length 2651;				
Best Local Similarity	57.3%;	Pred. No. 0.2;						
Matches	94;	Conservative	0;	Mismatches 70; Indels 0; Gaps 0				
QY	529 ATGAGGACACAGTGTGGCCGTGTGTGGGAGCGGCGGAGGCTCATCTGTGACGGGTGCC							
DB	164 ACGATGAGACACTGTGCGCGGCTCGACGAGGGGCCAAGCCTGACAGCCCTGCGGACCTGCC							
QY	589 CTCGGGCTTCCACCTGCGCTGCTGTCCCTTCGCTCGGAGATCCCGAGTGGACCT							
DB	244 CGGGGGCCCTACCACTCAGCTGCTGTGAGCCACCCCTCAAGAGCGGGCCCAAGGGCGTGT							
QY	649 GGAGGTGCTCCAGCTGCGCTGCAAGGCAACGATCCAGAGGTGCAG							
DB	304 GGGTGTGCCCGAGTGTCCAGGAAAGCGCTTAAAGAAACGAG							
RESULT 74								
ADB62798								
ID	ADB62798 standard; cDNA, 3498 BP.							
XX	ADB62798;							
XX	04-DEC-2003 (first entry)							
XX	Human cDNA encoding clone NT2RP70031340.							
XX	Human; ss; gene; pharmaceutical; diagnostic; gene therapy;							
KW	tissue regeneration; cell regeneration; membrane protein;							
KW	signal transduction-related protein; transcription-related protein;							
KW	osteoporosis; neurological disease; cancer; tumour.							
XX	Homo sapiens.							
OS	Homo sapiens.							
XX	Key							
FH	Location/Qualifiers							
FT	CDS							
FT	107..1576							
FT	/*tag= a							
FT	/product= "Clone NT2RP70031340 protein"							
XX	EPI308459-A2.							
XX	07-MAY-2003.							
XX	28-MAR-2002; 2002EP-00007401.							
XX	05-NOV-2001; 2001JP-00379298.							
XX	25-JAN-2002; 2002US-00350978.							
XX	(HELI-) HELIX RES INST.							
XX	(REAS-) RES ASSOC BIOTECHNOLOGY.							
XX	Iisogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,							
XX	Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,							
XX	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y,							
XX	MPI: 2003-450961/43.							
XX	P-PSDB; ADB64768.							
XX	New polynucleotides and polypeptides, useful for developing a diagnostic							
XX	marker or medicines for regulation of their expression and activity, or							
XX	as targets of gene therapy.							
XX	Claim 1; Page; 222pp; English.							
XX	The invention discloses a polynucleotide comprising a sequence selected							
XX	from 1970 fully defined nucleotide sequences which encode novel							
XX	polypeptides. Also claimed is a polypeptide encoded by the polynucleotide							
XX	or its partial peptide, an antibody binding to the polypeptide or peptide							
XX	of the polynucleotide, immunologically assaying the polypeptide or							
XX	peptide of the polynucleotide by contacting the polypeptide or peptide							
XX	with the antibody of the encoded protein, and observing the binding							
XX	between the two, a transformant carrying the polynucleotide in an							

CC	expressible manner and an antisense polynucleotide. The oligonucleotide
CC	is useful as a primer for synthesizing the polynucleotide, or as a probe
CC	for detecting the polynucleotide. The polynucleotides and encoded
CC	proteins are useful as pharmaceutical agents and many disease-related
CC	genes may be included in them, for developing a diagnostic marker or
CC	medicines for regulation of their expression and activity, or as targets
CC	of gene therapy. The genes are involved in tissue and/or cell
CC	regeneration. Membrane proteins, signal transduction-related proteins,
CC	transcription-related proteins, disease-related proteins and genes
CC	encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC	neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC	the activity or expression of the encoded protein to treat diseases. The
CC	sequence presented is a cDNA of the invention. Note: Some of the sequence
CC	data for this patent is not represented in the printed specification, but
CC	is based on sequence information supplied by the European Patent Office.
XX	
SQ	Sequence 3498 BP; 758 A; 1042 C; 1027 G; 671 T; 0 U; 0 Other;
QY	Query Match 3.4%; Score 52; DB 10; Length 3498;
Db	Best Local Similarity 57.3%; Pred. No. 0.2;
Matches	94; Conservative 0; Mismatches 70; Indels 0; Gaps 0
QY	529 ATGAGACGAGTGTGCGCTGTGTGTGCGGACGGCGGGAGCTATCTGCTGTGACGGCTGCC 588
Db	1032 ACGATGACACTGTGCGCCGCTGCAGACGAGGGGCCAACCTGCAAGCCTGCGGCACTTGCC 1091
QY	589 CTGGGGCCCTTCCACCTGTGCGCTGTGCTCCCTCGGCTCCGGAGATCCCAAGTGGACCT 648
Db	1092 CGGGGGCCTTACCACCTCACTGACGTGCTGCTGAGACCACTCTTAAGACGGCGCCCAAGGGCGTGT 1155
QY	649 GGAAGTGTCCAGCTGCTGTGCAAGCAACGATTCACGAGAGGTGACG 692
Db	1152 GGGTGTGCCCCAGGTGCCAGACGAAGGCGCTTAAGAAAGAACGAG 1195
RESULT 75	
ADB62384	
ID	ADB62384 standard; cDNA; 3584 BP.
XX	
AC	ADB62384;
XX	
DT	04-DEC-2003 (first entry)
XX	
DE	Human cDNA encoding clone FCBBF30142290.
XX	
KW	Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW	tissue regeneration; cell regeneration; membrane protein;
KW	signal transduction-related protein; transcription-related protein;
KW	osteoporosis; neurological disease; cancer; tumour.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	Location/Qualifiers
FT	CDS 679..1662
FT	/*tag= a
FT	/product= "Clone FCBBF30142290 protein"
XX	
PN	EP1308459-A2.
XX	
PD	07-MAY-2003.
XX	
PF	28-MAR-2002; 2002EP-00007401.
XX	
PR	05-NOV-2001; 2001JP-00379298.
XX	
PR	25-JAN-2002; 2002US-00350978.
XX	
PA	(HELI-) HELIX RES INST.
XX	(REAS-) RES ASSOC BIOTECHNOLOGY.
PI	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
PI	Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maseho Y,
XX	

DR MPI, 2003-450961/43.
DR P-PSDB; ADB64354.

XX New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.

XX Claim 1; Page; 222pp; English.

XX The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesizing the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours). The CDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a CDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.

XX SQ Sequence 3584 BP; 770 A; 1089 C; 1030 G; 695 T; 0 U; 0 Other;

Query Match 3.4%; Score 52; DB 10; Length 3584;
Best Local Similarity 57.3%; Pred. No. 0.21; Mismatches 0; Gaps 0;
Matches 94; Conservative 0; Indels 70; Indels 0; Gaps 0;

OY 529 ATGAGACGAGTGTGCGGTGTGCGGAGCGGCGGAGCTCATCTGTGACGGCTGCC 588
DB 1118 ACATATAGACACTGTGCGGCTGCGAGGAGGCGGCACTGAGCCCTGGGCACTGCC 1177

OY 589 CTCGGGCTTTCCACCTGCGGCTGCTCCCTCCGCTCCGGAGATCCCAATGGGAACT 648
DB 1178 CGGGGGCTTACCACTGCTGCTGAGAGCGGCCCTCTCAAGAGCGGCGCCCAAGGGCGTGT 1237

OY 649 GGAAGTGTCTCCAGCTGCGCTGCGGCAACAAGTCCAGGAGGTGCAG 692
DB 1238 GGGTGTGCCCCCAGGTGCGGAGGAGGCTTTAAAGAAAGCGAG 1281

RESULT 76
ADQ24519
ID ADQ24519 standard; DNA; 3710 BP.

XX ADQ24519;

XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 7339.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

PR 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Ziornik A;

XX MPI, 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.

XX Example 2; SEQ ID NO 7339; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual,
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

XX SQ Sequence 3710 BP; 784 A; 1129 C; 1032 G; 698 T; 0 U; 67 Other;

Query Match 3.4%; Score 52; DB 12; Length 3710;
Best Local Similarity 57.3%; Pred. No. 0.21; Mismatches 0; Gaps 0;
Matches 94; Conservative 0; Indels 70; Indels 0; Gaps 0;

OY 529 ATGAGACGAGTGTGCGGTGTGCGGAGCGGCGGAGCTCATCTGTGACGGCTGCC 588
DB 1203 ACATATAGACACTGTGCGGCTGCGAGGAGGCGGCACTGAGCCCTTGGCACTGCC 1262

OY 589 CTCGGGCTTTCCACCTGCGGCTGCTCCCTCCGCTCCGGAGATCCCAATGGGAACT 648
DB 1263 CGGGGGCTTACCACTGCTGCTGAGAGCGGCCCTCTCAAGAGCGGCGCCCAAGGGCGTGT 1322

OY 649 GGAAGTGTCTCCAGCTGCGCTGCGGCAACAAGTCCAGGAGGTGCAG 692
DB 1323 GGGTGTGCCCCCAGGTGCGGAGGAGGCTTTAAAGAAAGCGAG 1366

RESULT 77
ADA71938
ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 5263.

XX Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

CC preventing, and treating disorders associated with an abnormal expression
CC or activity of MAP (nucleic acid-associated proteins), such as
CC neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's
CC disease), muscular disorders (e.g. myotonic dystrophy, catatonias),
CC endocrine disorders (e.g. diabetes, Grave's disease), cancers (e.g.
CC leukemia, cervical or breast cancers), immunological disorders (e.g.
CC scleroderma, systemic lupus erythematosus, allergies), gastrointestinal
CC disorders (e.g. Crohn's disease), renal disorders (e.g. Goodpasture's
CC syndrome), infections (e.g. viral, bacterial, fungal, parasitic,
CC protozoal, helminthic), cardiovascular disorders (e.g. atherosclerosis),
CC or hepatic diseases (e.g. cirrhosis). NAAp or its fragments may also be
CC used in screening for compounds that specifically bind to and modulate
CC the activity of NAAp. The polynucleotides can be used to create humanised
CC animals or transgenic animals to model human disease.

XX
SQ Sequence 2226 BP; 702 A; 459 C; 583 G; 482 T; 0 U; 0 Other;

Query Match 3.2%; Score 48.8; DB 12; Length 2226;
Best Local Similarity 65.7%; Pred. No. 0.93; Mismatches 0; Gaps 0;
Matches 71; Conservative 0; Indels 0;

QY 515 CCTGTGCCAGAAATAGAGACGAGTGTGCGGTGTGCGGACGGGAGCTCATCTG 574
DB 1280 CCTGTATGAGAAACTTGATGAGTGTGCGGAGCGAGGAGCTGTCTG 1339
QY 575 CTGTACGGCTGCCCTCGGCGCTTCCACCTGCGCTGTGCTCCCTCC 622
DB 1340 TTGCGCACTGTGTTCAAGAGTCTTCCATGAGGAGCTGCACATCCACC 1387

RESULT 79
AAH16238
ID AAH16238 standard; cDNA; 3079 BP.

AC AAH16238;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15067.
XX
KW Human; primer; detection; diagnosis; antilease therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
FN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
XX
PR 27-AUG-1999; 99JP-00300253.
XX
PR 11-JAN-2000; 2000JP-00118776.
XX
PR 02-MAY-2000; 2000JP-00183767.
XX
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
PI WPI; 2001-318749/34.
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesising polynucleotides, particularly the 5602 full-
XX
PT length cDNAs defined in the specification, and for the detection and/or
XX
PT diagnosis of the abnormality of the proteins encoded by the full-length
XX
PT cDNA.

PS Claim 8; SEQ ID NO 15067; 2537bp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesising 5602 full-
XX
XX length cDNAs defined in the specification. Where a primer set comprises:
XX
XX (a) an oligo-dT primer and an oligonucleotide complementary to the

CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antilease therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

XX
SQ Sequence 3079 BP; 837 A; 744 C; 771 G; 727 T; 0 U; 0 Other;

Query Match 3.2%; Score 48.8; DB 4; Length 3079;
Best Local Similarity 65.7%; Pred. No. 0.97; Mismatches 37; Indels 0; Gaps 0;
Matches 71; Conservative 0; Indels 0;

QY 515 CCTGTGCCAGAAATAGAGACGAGTGTGCGGTGTGCGGACGGGAGCTCATCTG 574
DB 679 CCTGTATGAGAAACTTGATGAGTGTGCGGAGCGAGGAGCTGTCTG 738
QY 575 CTGTACGGCTGCCCTCGGCGCTTCCACCTGCGCTGTGCTCCCTCC 622
DB 739 TTGCGCACTGTGTTCAAGAGTCTTCCATGAGGAGCTGCACATCCACC 786

RESULT 80
AAK52998
ID AAK52998 standard; cDNA; 2171 BP.

AC AAK52998;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2527.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
FN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
XX
PR 27-APR-2000; 2000US-00560875.
XX
PR 20-JUN-2000; 2000US-00598075.
XX
PR 19-JUL-2000; 2000US-00620325.
XX
PR 01-SEP-2000; 2000US-00654936.
XX
PR 15-SEP-2000; 2000US-00663561.
XX
PR 20-OCT-2000; 2000US-00693325.
XX
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX
PI Ma Y, Zhao Q, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
XX
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

DR WPI: 2001-476283/51.
 DR P-PSDB; AAM79865.
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 PS Claim 1; Page 4806-4807; 6221pp; English.
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 2171 BP; 349 A; 759 C; 688 G; 375 T; 0 U; 0 Other;
 Query Match 3.1%; Score 48.6; DB 4; Length 2171;
 Best Local Similarity 52.2%; Pred. No. 1;
 Matches 108; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
 QY 543 GCCGTGTGCGGAGACGGGGAGCTCATCTGTGACGGCTGCCCTGGGCTTCCAC 602
 DB 91 GCCGTGTGCGGAGCTCATCTGTGACGGCTGCCCTGGGCTTCCATCGAT 150
 QY 603 CTGGCGCTGCTGCTCCCTCCGCTCCGCGAGATCCCAAGTGGACCTGGAGTGTCTCAGC 662
 DB 151 GCCGCCCGCAGCTCAAGCGCAGCCAGTGTCTCCGAGGCGCCGCTCGCCCGCCCGCCG 210
 QY 663 TGCCTGCAGGCAACAGTCCAGAGGTGACGCCCGGCGCAGAGAGCCCGCCCGCAGAG 722
 DB 211 GCCCTTAAGCACCCTGGCCCAAGAGACCTGGCGCGCAGCCGCAAGGGGCCCGCAGTG 270
 QY 723 CCACCCGTGGAGACCCCGCTCCCGCCG 749
 DB 271 CCGCCCCCGCAGGCGCCAGCCCGCAGCCG 297
 RESULT 81
 AAK52014
 ID AAK52014 standard; cDNA; 2179 BP.
 AC AAK52014;
 XX
 DT 06-NOV-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 559.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US004098.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich K;
 DR WPI: 2001-476283/51.
 DR P-PSDB; AAM78881.
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 PS Claim 1; Page 2022-2024; 6221pp; English.
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 2179 BP; 358 A; 757 C; 689 G; 375 T; 0 U; 0 Other;
 Query Match 3.1%; Score 48.6; DB 4; Length 2179;
 Best Local Similarity 52.2%; Pred. No. 1;
 Matches 108; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
 QY 543 GCCGTGTGCGGAGACGGGGAGCTCATCTGTGACGGCTGCCCTGGGCTTCCAC 602
 DB 91 GCCGTGTGCGGAGCTCATCTGTGACGGCTGCCCTGGGCTTCCATCGAT 150
 QY 603 CTGGCGCTGCTGCTCCCTCCGCTCCGCGAGATCCCAAGTGGACCTGGAGTGTCTCAGC 662
 DB 151 GCCGCCCGCAGCTCAAGCGCAGCCAGTGTCTCCGAGGCGCCGCTCGCCCGCCCGCCG 210
 QY 663 TGCCTGCAGGCAACAGTCCAGAGGTGACGCCCGGCGCAGAGAGCCCGCCCGCAGAG 722
 DB 211 GCCCTTAAGCACCCTGGCCCAAGAGACCTGGCGCGCAGCCGCAAGGGGCCCGCAGTG 270
 QY 723 CCACCCGTGGAGACCCCGCTCCCGCCG 749
 DB 271 CCGCCCCCGCAGGCGCCAGCCCGCAGCCG 297
 RESULT 82
 ADO35478/C
 ID ADO35478 standard; DNA; 3005 BP.
 AC ADO35478;
 XX
 DT 26-AUG-2004 (first entry)
 DE Novel mouse gene sequence #151.
 XX
 KW mouse; murine; cancer; psoriasis; ulcerative colitis; inflammation;
 KW ischaemic heart disease; thrombosis; immune disorder; bacterial disorder;
 KW viral disorder; ds; gene.
 XX
 OS Mus sp.
 XX
 PN WO2004046310-A2.

```

PD 03-JUN-2004.
XX
XX 24-OCT-2003; 2003WO-US033948.
XX
XX 15-NOV-2002; 2002US-0426916P.
XX 04-DEC-2002; 2002US-0431158P.
XX 05-DEC-2002; 2002US-043145P.
XX 05-DEC-2002; 2002US-043160P.
XX 09-JUN-2003; 2003US-0476621P.
XX 09-JUN-2003; 2003US-0476632P.
XX 08-JUL-2003; 2003US-0485217P.
XX 08-JUL-2003; 2003US-0485359P.
XX 08-AUG-2003; 2003US-0493332P.
XX 08-AUG-2003; 2003US-0493356P.
XX
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
XX Williams LT, Chu K, Lee E, Hestir K, Hayashizaki Y, Kamiya M;
XX WPI; 2004-431966/40.
XX
XX New mouse nucleic acid molecules and polypeptides, useful for treating
XX cancer, psoriasis, ulcerative colitis, inflammation, ischemic heart
XX disease or thrombosis.
XX
XX Claim 1, SEQ ID NO 151; 263pp; English.
XX
XX The invention comprises 744 novel mouse DNA sequences (genes). The DNA
XX sequences of the invention are useful for treating cancer, psoriasis,
XX ulcerative colitis, inflammation, ischemic heart disease, thrombosis,
XX immune disorders, bacterial disorders and viral disorders. The present
XX nucleic acid represents a mouse DNA sequence of the invention. NOTE: The
XX present DNA sequence is not shown in the specification, but has been
XX retrieved from the WIPO website.
XX
XX Sequence 3005 BP; 782 A; 632 C; 849 G; 742 T; 0 U; 0 Other;
SQ
Query Match 3.1%; Score 48; DB 12; Length 3005;
Best Local Similarity 48.2%; Pred. No. 1.4;
Matches 135; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
QY 1010 GCGCTCCACCTTCCGACCGGACCTCCCGGAGCGGCTGCGTGCAGATCTG 1069
DB 467 GAGCGCCATGTCGTGGTGGCCGCCACCCGCTCGGCTCGGCGCGGCTGCGCGCGG 408
QY 1070 CTCAGAGACGTGACCCCAAGCCCTTGAGGGGCTGTCGCCCCAGCCCGCCCT 1129
DB 407 GCTGGGTGACGAAGCTTCCGCTGCGGCGCGGAGCGCTCTAAGCCGAGCCCGGCA 348
QY 1130 GGGCCCTGGGCTGCCAAGATGACACTGCCAGTACAGACCCGCTCTGCACAGGATGA 1189
DB 347 GGTGCTCTCGCTTGTGGTGGCCGCCGACGCGCGCGCTTCAAGCCGAGCGGCGGA 288
QY 1190 CCGTAGTCCCTTCTGAGGAGACACCTTCATGATGATCTCTCAATGAGGCAATCCAG 1249
DB 287 GGAAGAGCTTCGCTTCAAGACATCTGAGGCGCGGAGCGCGCGCTGTCGCGGAG 228
QY 1250 CATGGCCGTCGCGCGCGCCCTTCCCTCTGACCCCAAG 1289
DB 227 CCTCGCGCTCTGAGGAGCTCGCGTGTCCCTCAGCCCAAG 188

```

RESULT 83

AA575442
ID AA575442 standard; cDNA; 924 BP.

AA575442;
XX

13-FEB-2002 (first entry)
XX

DNA encoding novel human diagnostic protein #11246.
DE

Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW

```

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX OS
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YF;
XX
XX WPI; 2001-639362/73.
XX P-PsDB; ABG11255.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1, SEQ ID NO 11246; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probe, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantifying a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AA564197-AA594554 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 924 BP; 259 A; 192 C; 264 G; 209 T; 0 U; 0 Other;
SQ
Query Match 3.1%; Score 47.8; DB 5; Length 924;
Best Local Similarity 67.7%; Pred. No. 1.3;
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 524 GAAGAATGAGACGAGTTCCTGTGTGCGGACGCGGAGACTATCTGCTGTACGG 583.
DB 51 GGAAGACTTGATGTGAGGTGCGGAGACGAGGAGAGCTGTCTGTTCGACAC 110
QY 584 CTGCGCTGGGCTTCACCTGAGCCCTGTCCTCCCTCC 622
DB 111 TTGTTCAAGATCTTTCATGAGGACTGCCAATCCACC 149

```

RESULT 84

AAH33830
ID AAH33830 standard; cDNA; 1006 BP.

AAH33830;
XX

03-SEP-2001 (first entry)
XX

Human colon cancer antigen encoding cDNA SEQ ID NO:886.
DE

```

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US026524.
XX 29-SEP-1999; 99US-0157137P.
XX 03-NOV-1999; 99US-0163280P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX MPI; 2001-235357/24.
XX P-PSDB; AAG74399.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX Claim 1; Page 2812; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cystostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patient's own production of P. Additionally, N may be used
XX to produce the colon cancer-associated Ps, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAH37196 to AAH37204 and AAB7789 represent
XX sequences used in the exemplification of the present invention. N.B.
XX Pages 666 to 682 and page 7053 of the sequence listing were missing at
XX time of publication, meaning no sequences are present for SEQ ID NO.1027
XX to 1052, 7921 and 7922
XX
XX Sequence 1006 BP; 339 A; 191 C; 232 G; 239 T; 0 U; 5 Other;
XX
XX Query Match 3.1%; Score 47.8; DB 4; Length 1006;
XX Best Local Similarity 60.3%; Pred. No. 1.4;
XX Matches 79; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
XX
XX 528 AATGAGACGAGTGTGCGGTGTCGGAGCGCGGAGCTATCTGCTGTGACGGCTGC 587
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 233 AATGAGACGAGTGTGCGGTGTCGGAGCGCGGAGCTATCTGCTGTGACGGAGATGT 292
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 588 CCTCGGGCCCTTCACCTGGCGCTGCTGCTCCCGCTCCGCGGAGATCCCACTGGAGCC 647
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 293 CCAAGAGCTTTTCATCTTAACCTTGTCACTTTCACACACTACTTAGCTTTCCAAAGTGGGAGC 352
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 648 TGGAGGTGCTC 658
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 353 TGGATATGCAC 363
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX RESULT 85
XX AAH16455
XX ID AAH16455 standard; cDNA; 2724 BP.
XX AC AAH16455;
XX DT 26-JUN-2001 (first entry)
XX

```

```

DE Human cDNA sequence SEQ ID NO:15457.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isegai T, Mshikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX MPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX Claim 8; SEQ ID NO 15457; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX oligonucleotides, all of which are used in the exemplification of the
XX present invention
XX
XX Sequence 2724 BP; 905 A; 602 C; 590 G; 627 T; 0 U; 0 Other;
XX
XX Query Match 3.1%; Score 47.8; DB 4; Length 2724;
XX Best Local Similarity 60.3%; Pred. No. 1.6;
XX Matches 79; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
XX
XX 528 AATGAGACGAGTGTGCGGTGTCGGAGCGCGGAGCTATCTGCTGTGACGGCTGC 587
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 1953 AATGAGACGAGTGTGCTGTGCGCAAAACGGAGGAGATCTGTGCTGCGCAAAATGT 2012
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 588 CCTCGGGCCCTTCACCTGGCGCTGCTGCTCCCGCTCCGCGGAGATCCCACTGGAGCC 647
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2013 CCAAGAGCTTTTCATCTTAACCTTGTCACTTTCACACACTACTTAGCTTTCCAAAGTGGGAGC 2072
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 648 TGGAGGTGCTC 658
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 2073 TGGATATGCAC 2083
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX

```

RESULT 86
ADP13566
ID ADP13566 standard; DNA; 3424 BP.
XX
AC ADP13566;
XX
DT 26-AUG-2004 (first entry)
XX
DE Renal cell carcinoma differentially expressed gene #302.
XX
KW de; diagnosis; non-blood disease; solid tumor; gene expression;
KM peripheral blood mononuclear cell; renal cell carcinoma; prostate cancer;
KM head/neck cancer; differential expression.
XX
OS Homo sapiens.
XX
PN WO2004048933-A2.
XX
PD 10-JUN-2004.
XX
PF 21-NOV-2003; 2003WO-US037481.
XX
PR 21-NOV-2002; 2002US-0427982P.
PR 03-APR-2003; 2003US-0459782P.
XX
PA (AMHP) WYETH.
PA (TWIN/) TWINE N C.
PA (BURC/) BURCZYNSKI M E.
PA (TREP/) TREPICCHIO W L.
PA (DORN/) DORNER A.
PA (STOV/) STOVER J A.
PA (SLON/) SLONI D K.
PI Twine NC, Burczynski ME, Trepicchio WL, Dorner A, Stover JA;
PI Sloni DK;
XX
DR WPI; 2004-460799/43.
XX
PT Diagnosing non-blood disease such as solid tumor, involves comparing
PT differential expression profile of specific genes in peripheral blood
PT sample of subject with reference expression profile of specific genes.
XX
PS Disclosure; SEQ ID NO 302; 350bp; English.
XX
CC The invention relate to a method of diagnosing (M1) non-blood disease
CC such as solid tumor by providing peripheral blood sample of human having
CC non-blood disease, and comparing an expression profile of specific genes
CC in the peripheral blood sample to reference expression profile of the
CC genes, where each of the genes is differentially expressed in peripheral
CC blood mononuclear cells (PBMCs) of patients having the disease as
CC compared to PBMCs of normal humans. The method is useful for diagnosing
CC non-blood disease such as solid tumor. The solid tumor is chosen from
CC renal cell carcinoma (RCC), prostate cancer and head/neck cancer. The
CC peripheral blood sample comprises enriched PBMCs. The peripheral blood
CC sample is a whole blood sample (claimed). (M1) is useful for identifying
CC genes that are differentially expressed in peripheral blood samples
CC isolated at different stages of progression, development or treatment of
CC RCC and/or other solid tumors. This sequence corresponds to a gene that
CC is differentially expressed and detected by the method of the invention.
CC (Note: this sequence is not given as part of the printed specification
CC but was obtained from WIPO in electronic format at
CC ftp.wipo/pub/published_pct_sequences).
XX
SQ Sequence 3424 BP; 1029 A; 811 C; 843 G; 741 T; 0 U; 0 Other;
XX
Query Match 3.1%; Score 47.8; DB 12; Length 3424;
Best Local Similarity 60.3%; Pred. No. 1.6;
Matches 79; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
XX
QY 528 AATGAGACGAGTGTCCGTGTGTCCGAGACGGCGGAGACTCATCTGTGTGACGGCTGC 587
DB 2669 AATGAAGACTGTGTCTGTCTGTCCAAAACGAGGAGATCTTGTGCTGCGAAAATGT 2728

QY 588 CCTCGGCGCTTCCACTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGAGCC 647
DB 2729 CCMAAGCTCTTCATCTAATCTGTCTGTCCAACTACTTACAGTTCCAAAGTGGAGC 2788
QY 648 TGGAGGTGCTC 658
DB 2789 TGGATATGAC 2799
XX
RESULT 87
ADB80996
ID ADB80996 standard; DNA; 3510 BP.
XX
AC ADB80996;
XX
DT 04-DEC-2003 (first entry)
XX
DE RING-SH complex related DNA, SEQ ID NO 70.
XX
KW RING-SH 3; Gag protein; Gag late domain; PI3K; actin; myosin; Hsp60;
KM Hep70; Hsp90; STAM1; STAM2A; VHS-UIW; GTPase; E2 enzyme; tsg101;
KM cullin; RING-SH; clathrin; vircide; vaccine; antiviral; retrovirus;
KM rhadovirus; filovirus; gene; ds.
XX
OS Unidentified.
XX
PN WO2003033646-A2.
XX
PD 24-APR-2003.
XX
PF 31-JUL-2002; 2002WO-US024589.
XX
PR 31-JUL-2001; 2001US-0308958P.
PR 09-NOV-2001; 2001US-0345846P.
XX
PA (PROT-) PROTEOLOGICS INC.
XX
PI Greener T, Moskowitz H, Reiss Y, Alroy I;
XX
DR WPI; 2003-393509/37.
XX
PT P-PSDB; ADB80955.
XX
PT New isolated protein complex comprising a RING-SH 3 polypeptide and
PT another polypeptide, useful for detecting cells infected with a virus,
PT or for treating viral disorders caused by retroviruses, rhadoviruses,
XX
PS Disclosure; Fig 69; 176bp; English.
XX
CC The invention relates to a novel isolated protein complex comprising a
CC RING-SH 3 polypeptide and a polypeptide selected from a Gag protein, a
CC Gag late domain, PI3K, actin, myosin, Hsp60, Hep70, STAM1, STAM2A,
CC STAM2B, VHS-UIW, a GTPase, an E2 enzyme, tsg101, a cullin, RING-SH, and a
CC clathrin. The novel protein complex has vircide activity and can be used
CC to treat disorders as part of a vaccine. The protein complex and
CC composition are useful for detecting cells infected with a virus, for
CC identifying agents having antiviral activity, and for treating viral
CC disorders caused by retroviruses, rhadoviruses, or filoviruses. This
CC polynucleotide represents a DNA sequence relating to a protein comprising
CC the RING-SH complex of the invention.
XX
SQ Sequence 3510 BP; 1044 A; 832 C; 886 G; 748 T; 0 U; 0 Other;
XX
Query Match 3.1%; Score 47.8; DB 9; Length 3510;
Best Local Similarity 60.3%; Pred. No. 1.6;
Matches 79; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
XX
QY 528 AATGAGACGAGTGTCCGTGTGTCCGAGACGGCGGAGACTCATCTGTGTGACGGCTGC 587
DB 2740 AATGAAGACTGTGTCTGTCTGTCCAAAACGAGGAGATCTTGTGCTGCGAAAATGT 2799
QY 588 CCTCGGCGCTTCCACTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGAGCC 647

Db 2800 CCAAGGCTTTCATTAATTGTCATGTTCCAGCACTTAAGCTTCCAGTGGGAC 2859
QY 648 TGGAGTGCTC 658
Db 2860 TGGATATGCAC 2870

RESULT 88
ID ADL22564 standard; cDNA; 6730 BP.
AC ADL22564;
XX
XX 20-MAY-2004 (first entry)
XX
XX Human disease detection and treatment (MDDT) cDNA - SEQ ID 13.
DE
XX disease detection; MDDT; antiarteriosclerotic; antipsoriatic; cytostatic;
XX cell signalling; arteriosclerosis; psoriasis; cancer; human; ss; gene.
XX
XX Homo sapiens.
OS
XX WO2003062379-A2.
XX
XX 31-JUL-2003.
XX
XX 14-JAN-2003; 2003WO-US001363.
XX
XX 17-JAN-2002; 2002US-0349413P.
XX
XX 17-JAN-2002; 2002US-0349946P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL,
XX Yu JY, Tuason O, Yap PE, Amshey SR, Dam TC, Liu TF, Gerstin EH,
XX Peralta CH, Lewis SA, Chen AJ, Marwaha R, Lan RX, Urashka ME,
XX Kristnam SR, Koliuru V, Panesar IS;
XX
XX MPI; 2003-853443/79.
XX
XX P-PSDB; ADL22668.
XX
XX New isolated disease detection and treatment polynucleotide for
XX diagnosing or treating conditions associated with cell signaling e.g.
XX arteriosclerosis, psoriasis, and cancer.
XX
XX Claim 1; SEQ ID NO 13; 411pp; English.
XX
XX The invention relates to a novel isolated disease detection and treatment
XX (MDDT) DNA polynucleotide. The polynucleotide of the invention
XX demonstrates antiarteriosclerotic, antipsoriatic and cytostatic
XX activities and may be useful in a composition for detecting the
XX expression of a disease detection and treatment molecule polynucleotide.
XX The molecules of the invention may be utilized to diagnose or treat
XX conditions, diseases or disorders associated with cell signalling, such
XX as arteriosclerosis, psoriasis and cancer. The current sequence is that
XX of a human MDDT cDNA of the invention.
XX
XX Sequence 6730 BP; 2123 A; 1184 C; 1206 G; 2217 T; 0 U; 0 Other;
SQ

Query Match 3.1%; Score 47.8; DB 11; Length 6730;
Best Local Similarity 60.3%; Pred. No. 1.8;
Matches 79; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 528 AATGAGAGAGAGTGTCCGTGTGTCGGAGCGCGGGAGCTCATCTGCTGTGACGCGCTGC 587
Db 1078 AATGAAGACTGTGTGCTGTGTCGCAAAACGAGAGAGATCTTGTGCTGCGAAATATG 1137
QY 588 CCGCGGCGCTTCCACCTGGCGCTGTCCCGCTCCGCGCTCCGGAGATCCCAATGGAGACC 647
Db 1138 CCAAGAGCTTTCATTAATTGTCATGTTCCAACTACTTAAGCTTTCCAGTGGGAC 1197
QY 648 TGGAGTGCTC 658

Db 1198 TGGATATGCAC 1208

RESULT 89
ID ABL06755 standard; cDNA; 3223 BP.
AC ABL06755;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 14747.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
XX
XX MPI; 2001-656860/75.
XX
XX P-PSDB; ABB62652.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 14747; 21pp + Sequence listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3223 BP; 793 A; 935 C; 890 G; 605 T; 0 U; 0 Other;
SQ

Query Match 3.1%; Score 47.6; DB 4; Length 3223;
Best Local Similarity 55.9%; Pred. No. 1.8;
Matches 114; Conservative 0; Mismatches 84; Indels 6; Gaps 1;

QY 526 AGAATGAGAGAGAGTGTCCGTGTGTCGGAGCGCGGGAGCTCATCTGCTGTGACGCT 585
Db 202 ACAACACACATATTTGTGACGCTCGAGAGAGGGGCAATGTGCTGTGCGATCGAT 261
QY 586 GCCCTCGGCGCTTCCACCTGGCGCTGTCCCGCTCCGCTCCGGAG-----ATCCCA 639
Db 262 GCCCTTCACGCTTTCACCTGCAATGCAATGATTCACCTTTGAGGAGAGACATACCA 321
QY 640 GTGGAGACTGAGAGTGTCTCCAGCTGCTGCAAGCAACATCCAGAGAGTGACGCCCGGG 699
Db 322 GTGGGCAAGTGGGTGGCCACAGCTGTCCGATGAGCAAGCTTCCAGCGCGGCGCTCT 381
QY 700 CAGAGAGAGCGCGGCGCCAGAGGC 723
Db 382 CATCCAGGCGCAGTTCGCTGAGC 405


```

RESULT 90
AAV24559
ID AAV24559 standard; DNA; 2905 BP.
XX
XX AAV24559;
AC
XX 17-SEP-1998 (first entry)
DT
XX Leukocyte specific protein, Sp140, coding sequence.
DE
XX Leukocyte specific protein, Sp140, coding sequence.
XX
XX Sp140; leukocyte specific protein; gene transcription regulator; therapy;
XX autoimmune disease; viral infection; cancer; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 107..2368
XX /tag a
XX /product= "Sp140"
XX
XX MO9814569-A1.
XX
XX 09-APR-1998.
XX
XX 02-OCT-1997; 97MO-US017715.
XX
XX 02-OCT-1996; 96US-0027347P.
XX
XX (BLOC/) BLOC D B.
XX (BLOC/) BLOC K D.
XX
XX Bloch DB, Bloch KD;
XX
XX MPI: 1998-286419/25.
XX P-PSDB; AAW57747.
XX
XX New isolated gene transcription regulator, Sp140 - used to develop
XX products for the diagnosis and treatment of auto-immune diseases, viral
XX infections or cancers.
XX
XX Claim 1, Page 54-57; 81pp; English.
XX
XX This sequence encodes the leukocyte specific protein, Sp140 of the
XX invention. The Sp140 polypeptides act as gene transcription regulators.
XX They can be used to develop products for use in the diagnosis and
XX treatment of autoimmune diseases such as primary biliary cirrhosis,
XX rheumatoid arthritis, systemic lupus erythematosus, Sjogren's syndrome,
XX scleroderma and multiple sclerosis, viral diseases including those caused
XX by herpes simplex virus, cytomegalovirus, HIV, hepatitis virus, human T-
XX cell leukemia virus-1 (HTLV-1) and adenovirus, and cancers including
XX leukaemia, particularly acute promyelocytic leukaemia, cancers of the
XX breast, ovary, prostate, bone, liver, pancreas or spleen, sarcomas and
XX melanomas
XX
XX Sequence 2905 BP; 874 A; 610 C; 760 G; 661 T; 0 U; 0 Other;
SQ
XX
XX Query Match 3.1%; Score 47.2; DB 2; Length 2905;
XX Best Local Similarity 64.8%; Pred. No. 2.1;
XX Matches 70; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 515 CCTGTGCCAGAAATGAGACGAGTGTCCGTGTTCGGAGCGCGGAGCTCATCTG 574
DB 1816 CCTGTATGAGAAACCTGATGATGTGATGTCCTGGAGCGGAGGAGCTGTTCTG 1875
QY 575 CTGTGACGCGTGCCTCGGAGCTTCGACGCGCTGCTGCTCCCTCC 622
DB 1876 TTGCGACACTTGTTCAGAGTCTTCATGAGACGTCTCATCTCCGCC 1923

```

RESULT 91
AA575444

```

ID AAS75444 standard; cDNA; 2911 BP.
XX
XX AAS75444;
AC
XX 13-FEB-2002 (first entry)
DT
XX DNA encoding novel human diagnostic protein #11248.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX MPI: 2001-639362/73.
XX P-PSDB; ABG11257.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnosis, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1; SEQ ID NO 11248; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostic as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnosis, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2911 BP; 878 A; 609 C; 759 G; 665 T; 0 U; 0 Other;
SQ
XX
XX Query Match 3.1%; Score 47.2; DB 5; Length 2911;
XX Best Local Similarity 64.8%; Pred. No. 2.1;
XX Matches 70; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 515 CCTGTGCCAGAAATGAGACGAGTGTCCGTGTTCGGAGCGCGGAGCTCATCTG 574
DB 1816 CCTGTATGAGAAACCTGATGATGTGATGTCCTGGAGCGGAGGAGCTGTTCTG 1875
QY 575 CTGTGACGCGTGCCTCGGAGCTTCGACCTGCGCTGCTGCTCCCTCC 622
DB 1876 TTGCGACACTTGTTCAGAGTCTTCATGAGACGTCTCATCTCCGCC 1923

```

RESULT 92

AA575443
ID AAS75443 standard; cDNA; 3252 BP.

AC AAS75443;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #11247.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

PI MPI; 2001-639362/73.

DR P-PSDB; ABG11256.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PT Claim 1; SEQ ID NO 11247; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3252 BP; 987 A; 678 C; 866 G; 721 T; 0 U; 0 Other;

XX Query Match 3.1%; Score 47.2; DB 5; Length 3252;

XX Best Local Similarity 64.8%; Pred. No. 2.1;

XX Matches 70; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 515 CCTGTCCCAAGATGATGAGACGAGTGTGCTGTCCGACGCGGAGCTCATCTG 574

DB 2168 CCTTTATATGAAACCTGATGATGTATGATGTGCTCCGACGAGGAGGAGCTTTCTG 2227

QY 575 CTGTGACGGCTGCTCTCGGACCTTCCACCTGAGCTCTGTCCCTCC 622

DB 2228 TTGCGACACTTGTTCAGAGTCTTCATAGAGACTGTCAATCCGCC 2275

RESULT 93

AA575445
ID AAS75445 standard; cDNA; 1210 BP.

AC AAS75445;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #11249.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

PI MPI; 2001-639362/73.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PT Claim 1; SEQ ID NO 11249; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1210 BP; 248 A; 321 C; 368 G; 273 T; 0 U; 0 Other;

XX Query Match 3.0%; Score 47; DB 5; Length 1210;

XX Best Local Similarity 57.9%; Pred. No. 2.1;

XX Matches 106; Conservative 0; Mismatches 70; Indels 7; Gaps 1;

QY 524 GAGAGATGAGACGAGTGTGCTGTCCGACGCGGAGGAGCTCATCTGCTGACGG 583

DB 588 GAGAAACCTGTGATGATGTGAGTGTGCTCCGACGAGGAGGAGCTTTGTGCGACAC 647

QY 584 CTGCCCTCGGAGCTTCCACCTGAGCTGTCCCTCC-----GCTCCGAGAGATCC 636

```

Db      648 TTGTTCAAGAGCTTTCATGAGACTGTACATCCCGCTGTGAAAGCTGAGAGGATCT 707
Qy      637 CCAATGGAGACCTGAGAGTGTCTCAGCTGCTGCAGGCAACAGTCCAGAGAGTGACAGCCCC 696
Db      708 TCCACTGTGCATGCGCAGCTCTCCAGCAGAGTGCGCGCTTGTGTATGATGGGAAAGAGC 767
Qy      697 GGG 699
Db      768 AGG 770

RESULT 94
AAS29014
ID      AAS29014 standard; DNA; 6610 BP.
XX
AC      AAS29014;
XX
DT      21-NOV-2001 (first entry)
XX
DE      Genomic sequence #20 encoding novel human uterine motility polypeptide.
XX
KW      Human; uterine motility-association disorder; uterus; pregnancy; labour;
XX      menstrual cycle; gene therapy; ds.
OS      Homo sapiens.
XX
PN      WO200155201-A1.
XX
PD      02-AUG-2001.
XX
PF      17-JAN-2001; 2001WO-US001317.
XX
PR      31-JAN-2000; 2000US-0179065P.
PR      04-FEB-2000; 2000US-0180628P.
PR      24-FEB-2000; 2000US-0184664P.
PR      02-MAR-2000; 2000US-0186350P.
PR      16-MAR-2000; 2000US-0189874P.
PR      17-MAR-2000; 2000US-0190076P.
PR      18-APR-2000; 2000US-0198123P.
PR      19-MAY-2000; 2000US-0205115P.
PR      07-JUN-2000; 2000US-0209467P.
PR      28-JUN-2000; 2000US-0214886P.
PR      30-JUN-2000; 2000US-0215135P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      14-JUL-2000; 2000US-0217496P.
PR      14-JUL-2000; 2000US-0218290P.
PR      26-JUL-2000; 2000US-0220963P.
PR      26-JUL-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224518P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225213P.
PR      14-AUG-2000; 2000US-0225214P.
PR      14-AUG-2000; 2000US-0225266P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225447P.
PR      14-AUG-2000; 2000US-0225577P.
PR      14-AUG-2000; 2000US-0225758P.
PR      14-AUG-2000; 2000US-0225759P.
PR      18-AUG-2000; 2000US-0226279P.
PR      22-AUG-2000; 2000US-0226681P.
PR      22-AUG-2000; 2000US-0226686P.
PR      23-AUG-2000; 2000US-0227182P.
PR      30-AUG-2000; 2000US-0227009P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      01-SEP-2000; 2000US-0229345P.
PR      14-AUG-2000; 2000US-0229345P.
PR      05-SEP-2000; 2000US-0229509P.
PR      05-SEP-2000; 2000US-0229513P.
PR      06-SEP-2000; 2000US-0230437P.
PR      06-SEP-2000; 2000US-0230438P.
PR      08-SEP-2000; 2000US-0231242P.
PR      08-SEP-2000; 2000US-0231243P.
PR      08-SEP-2000; 2000US-0231244P.
PR      08-SEP-2000; 2000US-0231413P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0232080P.
PR      08-SEP-2000; 2000US-0232081P.
PR      12-SEP-2000; 2000US-0231968P.
PR      14-SEP-2000; 2000US-0232397P.
PR      14-SEP-2000; 2000US-0232398P.
PR      14-SEP-2000; 2000US-0232399P.
PR      14-SEP-2000; 2000US-0232400P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0233063P.
PR      14-SEP-2000; 2000US-0233064P.
PR      14-SEP-2000; 2000US-0233065P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      25-SEP-2000; 2000US-0234997P.
PR      25-SEP-2000; 2000US-0234998P.
PR      26-SEP-2000; 2000US-0235484P.
PR      27-SEP-2000; 2000US-0235834P.
PR      27-SEP-2000; 2000US-0235836P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      29-SEP-2000; 2000US-0236370P.
PR      02-OCT-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      02-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239935P.
PR      13-OCT-2000; 2000US-0239937P.
PR      20-OCT-2000; 2000US-0240960P.
PR      20-OCT-2000; 2000US-0241221P.
PR      20-OCT-2000; 2000US-0241785P.
PR      20-OCT-2000; 2000US-0241786P.
PR      20-OCT-2000; 2000US-0241787P.
PR      20-OCT-2000; 2000US-0241808P.
PR      20-OCT-2000; 2000US-0241809P.
PR      20-OCT-2000; 2000US-0241826P.
PR      01-NOV-2000; 2000US-0244617P.
PR      08-NOV-2000; 2000US-0246474P.
PR      08-NOV-2000; 2000US-0246475P.
PR      08-NOV-2000; 2000US-0246476P.
PR      08-NOV-2000; 2000US-0246477P.
PR      08-NOV-2000; 2000US-0246478P.
PR      08-NOV-2000; 2000US-0246523P.
PR      08-NOV-2000; 2000US-0246524P.
PR      08-NOV-2000; 2000US-0246525P.
PR      08-NOV-2000; 2000US-0246526P.
PR      08-NOV-2000; 2000US-0246527P.
PR      08-NOV-2000; 2000US-0246528P.
PR      08-NOV-2000; 2000US-0246532P.
PR      08-NOV-2000; 2000US-0246609P.
PR      08-NOV-2000; 2000US-0246610P.
PR      08-NOV-2000; 2000US-0246611P.
PR      08-NOV-2000; 2000US-0246613P.
PR      17-NOV-2000; 2000US-0249207P.
PR      17-NOV-2000; 2000US-0249208P.
PR      17-NOV-2000; 2000US-0249209P.
PR      17-NOV-2000; 2000US-0249210P.
PR      17-NOV-2000; 2000US-0249211P.
PR      17-NOV-2000; 2000US-0249212P.
PR      17-NOV-2000; 2000US-0249213P.
PR      17-NOV-2000; 2000US-0249214P.
PR      17-NOV-2000; 2000US-0249215P.
```


DE Human genomic DNA encoding partial novel secreted protein, Seq ID 303.
XX
KW Human; immunosuppressive; antiarthritic; ds; antineumatic; cytostatic;
KW cariant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmological; vulnary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disease; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155441-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001320.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214866P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.

PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241825P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.

PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233065P.
PR 14-SEP-2000; 2000US-0233066P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476222/51.
XX
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, for treating blood clotting disorder,
PT hemophilia.
XX
PS Disclosure; SEQ ID NO 304; 601bp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodysplasia,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence is a genomic DNA encoding a partial novel secreted protein of
CC the invention. Note: The sequence data for this patent did not form part
Query Match 3.0%; Score 47; DB 4; Length 6610;
Best Local Similarity 53.6%; Pred. No. 2.6;
Matches 143; Conservative 0; Mismatches 120; Indels 4; Gaps 2;
QY 946 GCGGGGTGCGGAGATGATGATGAGACGTGCGGTGATGATGATGCGCGCTTCC 1005
DB 7 GCTGTGAGTGGCGGAGGATGATGATGCGCGCGCTGCGATGATGATGATGATGCTTCC 66
QY 1006 ACTGCGCTGCTCCACTTCCAGCGCGACCTCCCGGCGGAGACGGGCTGCGCTTCAGAT 1065
DB 67 GCGGCGGCTGCGCACCTGTGAGACGCGATGATGCTGCGAGCCGGGCTTACCGGCAAGT 126
QY 1066 CTGCTCAGAGAGCTGATCCCAAGCGCCCTGTGAGAGGGGTGTGAGCCCCCGGCCCC 1125
DB 127 ACTG-TGCGAGACCGTGGCCCGCGGCTTCTTACGAGCTTGGGCTGTGCGCG--CCGGTAA 182


```
OY      1126 GCCTGGCCCTTGCGCTGCGAGATGACACTGCCAGTACGAGCCCGCTTGACACAGG 1185
DB      183 GCGCGCTCTCTCGCGAGGAGAGAGAGTGGAGGCGGCGCGCGCGATGGGAGACAGG 242
OY      1186 ATGACCTGGAGTCCCTTCTGAGGAGC 1212
DB      243 GTGAGCGGGGTCCACGCTTCCCTTACG 269

RESULT 98
ABA06837
ID      ABA06837 strand: DNA; 6610 BP.
XX      ABA06837;
AC      10-JAN-2002 (first entry)
XX      Human genomic DNA SEQ ID NO: 923.
DE      Human; gene therapy; neural disorder; immune system disorder;
XX      muscular disorder; reproductive disorder; gastrointestinal disorder;
XX      pulmonary disorder; cardiovascular disorder; renal disorder;
XX      proliferative disorder; inflammation; ds.
OS      Homo sapiens.
XX      WO200154474-A2.
XX      02-AUG-2001.
XX      17-JAN-2001; 2001WO-US001349.
XX      31-JAN-2000; 2000US-0179065P.
PR      04-FEB-2000; 2000US-0180628P.
PR      24-FEB-2000; 2000US-0184664P.
PR      02-MAR-2000; 2000US-0186350P.
PR      16-MAR-2000; 2000US-0189874P.
PR      17-MAR-2000; 2000US-0190076P.
PR      18-APR-2000; 2000US-0198123P.
PR      19-MAY-2000; 2000US-0205151P.
PR      07-JUN-2000; 2000US-0209467P.
PR      28-JUN-2000; 2000US-0214886P.
PR      30-JUN-2000; 2000US-0215135P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      14-JUL-2000; 2000US-0217969P.
PR      26-JUL-2000; 2000US-0220963P.
PR      26-JUL-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224518P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225213P.
PR      14-AUG-2000; 2000US-0225214P.
PR      14-AUG-2000; 2000US-0225266P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225447P.
PR      14-AUG-2000; 2000US-0225477P.
PR      14-AUG-2000; 2000US-0225577P.
PR      14-AUG-2000; 2000US-0225758P.
PR      14-AUG-2000; 2000US-0225759P.
PR      18-AUG-2000; 2000US-0226279P.
PR      22-AUG-2000; 2000US-0226681P.
PR      22-AUG-2000; 2000US-0226688P.
PR      23-AUG-2000; 2000US-0227182P.
PR      23-AUG-2000; 2000US-0227009P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      01-SEP-2000; 2000US-0229345P.
PR      05-SEP-2000; 2000US-0229509P.
PR      05-SEP-2000; 2000US-0229513P.
PR      06-SEP-2000; 2000US-0230437P.
PR      06-SEP-2000; 2000US-0230438P.
PR      08-SEP-2000; 2000US-0231242P.
PR      08-SEP-2000; 2000US-0231243P.
PR      08-SEP-2000; 2000US-0231244P.
PR      08-SEP-2000; 2000US-0231413P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0232080P.
PR      08-SEP-2000; 2000US-0232081P.
PR      12-SEP-2000; 2000US-0231968P.
PR      14-SEP-2000; 2000US-0232397P.
PR      14-SEP-2000; 2000US-0232398P.
PR      14-SEP-2000; 2000US-0232399P.
PR      14-SEP-2000; 2000US-0232400P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0233063P.
PR      14-SEP-2000; 2000US-0233064P.
PR      14-SEP-2000; 2000US-0233065P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      25-SEP-2000; 2000US-0234997P.
PR      25-SEP-2000; 2000US-0234998P.
PR      26-SEP-2000; 2000US-0235484P.
PR      27-SEP-2000; 2000US-0235834P.
PR      27-SEP-2000; 2000US-0235836P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      29-SEP-2000; 2000US-0236370P.
PR      02-OCT-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      02-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239935P.
PR      13-OCT-2000; 2000US-0239937P.
PR      20-OCT-2000; 2000US-0241221P.
PR      20-OCT-2000; 2000US-0241785P.
PR      20-OCT-2000; 2000US-0241786P.
PR      20-OCT-2000; 2000US-0241787P.
PR      20-OCT-2000; 2000US-0241808P.
PR      20-OCT-2000; 2000US-0241809P.
PR      20-OCT-2000; 2000US-0241826P.
PR      01-NOV-2000; 2000US-0244617P.
PR      08-NOV-2000; 2000US-0244674P.
PR      08-NOV-2000; 2000US-0246475P.
PR      08-NOV-2000; 2000US-0246476P.
PR      08-NOV-2000; 2000US-0246477P.
PR      08-NOV-2000; 2000US-0246478P.
PR      08-NOV-2000; 2000US-0246523P.
PR      08-NOV-2000; 2000US-0246524P.
PR      08-NOV-2000; 2000US-0246525P.
PR      08-NOV-2000; 2000US-0246526P.
PR      08-NOV-2000; 2000US-0246527P.
PR      08-NOV-2000; 2000US-0246528P.
PR      08-NOV-2000; 2000US-0246532P.
PR      08-NOV-2000; 2000US-0246609P.
PR      08-NOV-2000; 2000US-0246610P.
PR      08-NOV-2000; 2000US-0246611P.
PR      08-NOV-2000; 2000US-0246613P.
PR      17-NOV-2000; 2000US-0249207P.
PR      17-NOV-2000; 2000US-0249208P.
PR      17-NOV-2000; 2000US-0249209P.
PR      17-NOV-2000; 2000US-0249210P.
PR      17-NOV-2000; 2000US-0249211P.
PR      17-NOV-2000; 2000US-0249212P.
PR      17-NOV-2000; 2000US-0249213P.
PR      17-NOV-2000; 2000US-0249214P.
PR      17-NOV-2000; 2000US-0249215P.
```

PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0253678P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-476161/51.
XX
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
XX condition.
XX
PS Disclosure; SEQ ID NO 923; 859pp + Sequence Listing; English.
XX
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC cDNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a genomic DNA of the invention
XX
SQ Sequence 6610 BP; 1208 A; 2130 C; 2123 G; 1149 T; 0 U; 0 Other;
Query Match 3 0%; Score 47; DB 4; Length 6610;
Best Local Similarity 53.6%; Pred. No. 2.6; Mismatches 120; Indels 4; Gaps 2;
Matches 143; Conservative 0;
QY 946 GCGGGGTTGCGAGATGTAGTACGACGTGCTCGGTGTACTGACGCGGCTTCC 1005
DB 7 GCTGTACAGTGGCGGAGCGTACGTTGCGGGCGCGTGGCATGCTACTGCGACGTGCTTC 66
QY 1006 ACTGGGCTGTCCACTTCCAGCCGGCACTCCCGGCGGAGCGGCGCTTGCCTGCAAT 1065
DB 67 GGGGGCGCTGCACCCCTGTGACGCGACGTGCTGCTGCGAGCGCGGCTTACCGCGCAAGT 126
QY 1066 CTTGCTCAGAGAGAGTACCCAGCCCGCTGTGAGAGGGGTGTGGCCCGCCAGCCCGGCC 1125
DB 127 ACTG-TTCGGAGACCGTGGCCCGCGGCTTCTACGGCTTGGGCTGTGCGCG---CCGGTAA 182
QY 1126 GCGTGGCCCTTGCGGCGCTGCGCAAGATGACACTGTCCAGTACGAGCCGCTTGCACAGG 1185
DB 183 GCGCGGTCTCTCGGCGCAGGAGGAGAGTGAAGGAGGCGGCGCGCATGAGGAGACAGG 242
QY 1186 ATGACCTGAGATCCCTTCTGAGCGAGC 1212
DB 243 GTGAGCGGGTCCAGGCTTCCCTTACG 269

AC ABA06838;
XX 10-JAN-2002 (first entry)
XX
XX
DE Human genomic DNA SEQ ID NO: 924.
XX
XX
KW Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ds.
OS Homo sapiens.
XX
XX
PN WO200154474-A2.
PD 02-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US001349.
XX
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227109P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.

```

PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234232P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240360P.
PR 20-OCT-2000; 2000US-0241211P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244647P.
PR 08-NOV-2000; 2000US-02446475P.
PR 08-NOV-2000; 2000US-02446476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249397P.
PR 17-NOV-2000; 2000US-0249399P.
PR 17-NOV-2000; 2000US-0249399P.
PR 17-NOV-2000; 2000US-0249399P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476161/51.
XX
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition.
XX
XX
PS Disclosure; SEQ ID NO 924; 859pp + Sequence Listing; English.
XX
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNA. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a genomic DNA of the invention
XX
SQ Sequence 6610 BP; 1208 A; 2130 C; 2123 G; 1149 T; 0 U; 0 Other;

Query Match 3.0%; Score 47; DB 4; Length 6610;
Best Local Similarity 53.6%; Pred. No. 2.6; Indels 4; Gaps 2;
Matches 143; Conservative 0; Mismatches 120;

QY 946 GCGGGGTGCGGAGATGTAGACGCTGCGGTGCTACTGACGCGCTGCTTCC 1005
DB 7 GCTGTCAAGTCCCGCAGCGCTTACGCTTGGCGCGCGCTGCGCTACTGCTGCTTCC 66
QY 1006 ACTGGCGCTGACCACTTCCAGCGCACTCCCGCGCGGACGCGCGCTGCGCTGCAAT 1065
DB 67 GCGGCGGCTGACCACTTGGAGCGGCACTGCTGCGCGCGCTGCGCGCGCTGCGCGCAAT 126
QY 1066 CCTGCTCAGGAGACGTCAGCCAGCCCGCTGTGAGGCGGTGTGCGCGCGCGCGCGCC 1125
DB 127 ACTG-TCGCGAGCGCGCGCGCGCGCGCTTACGCTTGGCGCTGCTGCGCGCGCGGTA 182
QY 1126 GCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1185
DB 183 GCGCGGTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 242
QY 1186 ATGACCTGAGTCCCTTCTGAGCGAGC 1212
DB 243 GTGAGGCGGCGTCCAGCGCTTCCCTAGC 269

RESULT 100
ABT07848
ID ABT07848 standard; DNA; 6610 BP.
XX
AC ABT07848;
XX
DT 14-NOV-2002 (first entry)
XX
DE Novel human nucleic acid SEQ ID No 147.
XX
XX Immunostimulant; antirheumatic; antiarthritic; neuroprotective;
XX anti-allergic; antidiabetic; antiasthmatic; antiinflammatory; nootropic;
XX immunosuppressive; anticoagulant; chromolytic; antithrombotic;
XX cytostatic; nephroprotective; antiparkinsonian; gynecological; virucide;
XX antibacterial; antihypertensive; fungicide; HGPAT05; HMAEB95; HTNBK01;
XX immunodeficiency; autoimmune disorder; allergic reaction; cardiovascular;
XX inflammatory condition; graft-versus-host disease; reproductive system;
XX blood-related disorder; hyperproliferative; endocrine; neurological;

```

KM respiratory; renal; infectious disease; gastrointestinal; gene therapy;
KM neuronal growth; neuronal disorder; neuro-degenerative condition;
KM keratinocyte growth; human; ds.
XX Homo sapiens.
XX OS
XX US2002086330-A1.
XX PD
XX 04-JUL-2002.
XX
PF 17-JAN-2001; 2001US-00764893.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 28-JUN-2000; 2000US-0214886P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 14-JUL-2000; 2000US-0217496P.
XX 26-JUL-2000; 2000US-0220863P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 22-AUG-2000; 2000US-0226688P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 08-SEP-2000; 2000US-0229513P.
XX 08-SEP-2000; 2000US-0231413P.
XX 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234977P.
XX 27-SEP-2000; 2000US-0235834P.
XX 29-SEP-2000; 2000US-0236327P.
XX 29-SEP-2000; 2000US-0236367P.
XX 29-SEP-2000; 2000US-0236368P.
XX 29-SEP-2000; 2000US-0236369P.
XX 29-SEP-2000; 2000US-0236370P.
XX 02-OCT-2000; 2000US-0236802P.
XX 02-OCT-2000; 2000US-0237037P.
XX 02-OCT-2000; 2000US-0237038P.
XX 02-OCT-2000; 2000US-0237039P.
XX 02-OCT-2000; 2000US-0237040P.
XX 13-OCT-2000; 2000US-0239335P.
XX 20-OCT-2000; 2000US-0240960P.
XX 20-OCT-2000; 2000US-0241785P.
XX 20-OCT-2000; 2000US-0241809P.
XX 01-NOV-2000; 2000US-0244617P.
XX 17-NOV-2000; 2000US-0249299P.
XX 08-DEC-2000; 2000US-0251856P.
XX 08-DEC-2000; 2000US-0251868P.
XX 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN S A.
XX (RUBE/) RUBEN S M.
XX (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX MPI; 2002-665432/71.
XX
XX Novel polypeptide useful for diagnosis, prognosis, prevention, and
XX treatment of immune, hyperproliferative, renal, respiratory, and
XX

PT cardiovascular, reproductive, endocrine, gastrointestinal and
PT neurological disorders.

PS Disclosure, Page 323-326; 335pp; English.

XX The invention relates to an isolated polypeptide comprising a sequence at
XX least 90% identical to a full length protein sequence selected from 55
XX CC sequences given in the specification such as a sequence of 163, 74 or 140
XX CC amino acids fully defined in the specification, or the encoding sequence
XX CC contained in 49 cDNA clones given in specification e.g. HCFAT05, HMAAB35
XX CC or HTNEM01. The protein and its encoding nucleic acid are useful for
XX CC diagnosing a pathological condition or susceptibility to a pathological
XX CC condition in a subject and for preventing, treating or ameliorating a
XX CC medical condition. The protein, its encoding nucleic acid and an isolated
XX CC antibody that can bind to the protein are useful in treating, preventing,
XX CC diagnosing and/or prognosing immunodeficiencies, autoimmune disorders,
XX CC allergic reactions and conditions, inflammatory conditions, graft-versus-
XX CC host disease, blood-related disorders, hyperproliferative disorders,
XX CC renal disorders, cardiovascular disorders, respiratory disorders,
XX CC neurological disorders, endocrine disorders, reproductive system
XX CC disorders, infectious diseases, and gastrointestinal disorders. The
XX CC protein of the invention is useful to stimulate neuronal growth and to
XX CC treat, prevent, and/or diagnose neuronal damage which occurs in certain
XX CC neuronal disorders or neuro-degenerative conditions, for stimulating
XX CC keratinocyte growth, to prevent hair loss, to modulate mammalian
XX CC characteristics such as body height, weight, hair color, and to increase
XX CC or decrease storage capabilities, fat content, lipid, protein,
XX CC carbohydrate, vitamins, minerals, cofactors or other nutritional
XX CC components. The nucleic acid of the invention can be used in gene
XX CC therapy. This polynucleotide sequence represents one of the novel nucleic
XX acids of the invention

SQ Sequence 6610 BP; 1208 A; 2130 C; 2123 G; 1149 T; 0 U; 0 Other;

Query Match 3.0%; Score 47; DB 6; Length 6610;

Best Local Similarity 53.6%; Pred. No. 2.6;

Matches 143; Conservative 0; Mismatches 120; Indels 4; Gaps 2;

QY 946 GCGGGGTGGGGAATGATACGACGTCGCGGTGTAACATGACGCGCGCTGCCCTCC 1005
DB 7 GCTGTCAAGTCCGCGAGCGTACGTTCCGCGCGCGCTGCGATGCTACTGCAAGTCTTCC 66
QY 1006 ACTGCGCTGTCACCTTCCAGCCGACCTCCCGGCCGCGGAGCGGAGCTGCGCTCAGAT 1065
DB 67 GCGGCGGCTGCGACCTGTGAGCGGCACTGTGCTGGAGCGGAGCTTACCGGCAAGT 126
QY 1066 CTGCTCAGAGAGCTGACCCCAAGCCCTGTGAGAGGGGTGCTGGCCCCCAAGCCCCGCC 1125
DB 127 ACTG-TCCGAGCCCTGCCCCCGCGCTTCTACGCTTGGGCTGTGCGCG--CCGATAA 182
QY 1126 GCGTGGCCCTGGGCGCTGCGCAAGATGACATGCGCAGTACAGAGCCGCTGTGACAGGG 1185
DB 183 GCGGCGTCTCTGCGGCAAGAGAGAGAGTGAAGAGGCGGCGGCCCGATGAGGAGACAGGG 242
QY 1186 ATGACCTGAGTCCCTTCTGAGCGAGC 1212
DB 243 GTGAGCGGGGTCCAGCCTTCCCTAGC 269

Search completed: February 9, 2005, 14:33:16
Job time : 768 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 09:10:56 ; Search time 157 Seconds
(without alignments)

6994.706 Million cell updates/sec

Title: US-09-508-658A-3

Perfect score: 1545
Sequence: 1 agagaagaagcagagctcttc.....aaaataataaataagctcg 1545

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 200 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCTUS COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfillseq.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match	Query Length	DB ID	Description
1	60	3.9	6328	3	US-08-913-832A-1 Sequence 1, Appl1
2	60	3.9	6328	4	US-09-249-181A-1 Sequence 1, Appl1
3	60	3.9	6328	4	US-09-158-707-1 Sequence 1, Appl1
4	60	3.9	6475	4	US-09-620-3120-325 Sequence 325, App
5	59.8	3.9	1053	4	US-07-945-295-1 Sequence 1, Appl1
6	59.8	3.9	1053	5	PCT-US91-06418-2 Sequence 2, Appl1
7	47.2	3.1	2905	3	US-08-942-686-1 Sequence 1, Appl1
8	44.6	2.9	2237	4	US-09-976-594-1064 Sequence 1064, Ap
9	44.6	2.9	4616	1	US-08-340-203A-1 Sequence 1, Appl1
10	44.6	2.9	4616	2	US-08-452-567-1 Sequence 1, Appl1
11	44.6	2.9	4616	2	US-08-452-427-1 Sequence 1, Appl1
12	44.6	2.9	4616	3	US-09-085-407-1 Sequence 1, Appl1
13	42.6	2.8	882	4	US-09-920-923B-36 Sequence 36, Appl1
14	42.6	2.8	960	4	US-09-526-400-3 Sequence 3, Appl1
15	42.6	2.8	1182	4	US-09-526-400-1 Sequence 1, Appl1
16	42.6	2.8	1469	5	PCT-US92-08258-1 Sequence 1, Appl1
17	42.2	2.7	3096	4	US-09-270-767-13712 Sequence 13712, A
18	42.2	2.7	4403765	3	US-09-103-840A-2 Sequence 1, Appl1
19	42.2	2.7	4411529	3	US-09-387-286-60 Sequence 60, Appl1
20	42	2.7	3612	4	PCT-US91-07635-5 Sequence 5, Appl1
21	41.6	2.7	1941	5	US-08-332-463-14 Sequence 14, Appl1
22	41.4	2.7	7218	1	US-09-312-762A-3 Sequence 3, Appl1
23	41.2	2.7	14707	4	US-09-511-974A-34 Sequence 34, Appl1
24	40.8	2.6	364	4	US-09-621-976-17202 Sequence 17202, A
25	40.6	2.6	430	4	US-09-621-976-16656 Sequence 16656, A
26	40.4	2.6	3033	4	US-09-724-797-81 Sequence 81, Appl1
27	40.4	2.6	3033	4	US-09-724-797-81 Sequence 81, Appl1

28	40.2	2.6	1556	2	US-08-881-857-1 Sequence 1, Appl1
29	40.2	2.6	1556	3	US-09-233-342A-1 Sequence 1, Appl1
30	40.2	2.6	1582	4	US-09-620-312D-389 Sequence 389, App
31	40.2	2.6	2399	2	US-08-969-106-9 Sequence 9, Appl1
32	40.2	2.6	2399	4	US-09-338-125-9 Sequence 9, Appl1
33	40	2.6	293	4	US-09-313-294A-604 Sequence 604, App
34	40	2.6	762	4	US-09-252-991A-11986 Sequence 11986, A
35	40	2.6	1125	4	US-09-252-991A-11986 Sequence 11986, A
36	39.6	2.6	954	3	US-08-680-506-5 Sequence 5, Appl1
37	39.6	2.6	1547	3	US-08-680-506-8 Sequence 8, Appl1
38	39.6	2.6	1587	3	US-08-680-506-6 Sequence 6, Appl1
39	39.6	2.6	2233	3	US-08-680-506-4 Sequence 4, Appl1
40	39.6	2.6	4184	2	US-08-785-310A-4 Sequence 2, Appl1
41	39.6	2.6	5027	3	US-08-680-506-2 Sequence 2, Appl1
42	39.6	2.6	47981	4	US-09-679-279-1 Sequence 1, Appl1
43	39.4	2.6	675	3	US-08-998-416-723 Sequence 723, App
44	39.2	2.5	1596	3	US-09-655-270A-2 Sequence 2, Appl1
45	39.2	2.5	1596	3	US-09-651-941-2 Sequence 2, Appl1
46	39.2	2.5	1596	4	US-09-655-597-2 Sequence 2, Appl1
47	39.2	2.5	12508	3	US-09-655-270A-1 Sequence 1, Appl1
48	39.2	2.5	12523	3	US-09-651-941-1 Sequence 1, Appl1
49	39.2	2.5	12523	4	US-09-655-597-1 Sequence 1, Appl1
50	39	2.5	531	4	US-09-252-991A-11993 Sequence 11993, A
51	39	2.5	564	4	US-09-252-991A-11993 Sequence 11993, A
52	39	2.5	603	4	US-09-252-991A-11962 Sequence 11962, A
53	39	2.5	645	4	US-09-252-991A-14041 Sequence 14041, A
54	38.8	2.5	808	4	US-09-252-991A-13744 Sequence 13744, A
55	38.8	2.5	704	4	US-09-252-991A-13605 Sequence 13605, A
56	38.8	2.5	984	4	US-09-252-991A-13494 Sequence 13494, A
57	38.8	2.5	1104	4	US-09-252-991A-1337 Sequence 1337, Ap
58	38.8	2.5	1275	4	US-09-252-991A-4071 Sequence 4071, Ap
59	38.8	2.5	2481	4	US-09-894-998A-35 Sequence 35, Appl1
60	38.6	2.5	11958	3	US-09-134-248-8 Sequence 8, Appl1
61	38.6	2.5	229354	4	US-09-765-400-64 Sequence 64, Appl1
62	38.6	2.5	229354	4	US-09-705-400-64 Sequence 64, Appl1
63	38.4	2.5	1683	4	US-09-724-797-21 Sequence 21, Appl1
64	38.4	2.5	3524	4	US-09-077-940A-3 Sequence 3, Appl1
65	38.4	2.5	3721	4	US-10-140-002-543 Sequence 543, App
66	38.4	2.5	4480	3	US-09-191-171-7 Sequence 7, Appl1
67	38.4	2.5	4480	3	US-09-385-707-7 Sequence 7, Appl1
68	38.2	2.5	6238	4	US-09-639-696C-6 Sequence 6, Appl1
69	38.2	2.5	1260	4	US-09-252-991A-6674 Sequence 6674, Ap
70	38.2	2.5	2163	4	US-09-252-991A-6641 Sequence 6641, Ap
71	38.2	2.5	2889	1	US-08-289-112-1 Sequence 1, Appl1
72	38.2	2.5	4403765	3	US-09-103-840A-2 Sequence 2, Appl1
73	38	2.5	702	1	US-08-458-568A-3 Sequence 3, Appl1
74	38	2.5	1335	5	PCT-US91-06532-1 Sequence 1, Appl1
75	38	2.5	2634	1	US-08-196-218-31 Sequence 31, Appl1
76	38	2.5	2634	1	US-08-681-953-31 Sequence 31, Appl1
77	38	2.5	2846	3	US-09-613-182-5 Sequence 5, Appl1
78	38	2.5	12001	1	US-08-458-568A-11 Sequence 11, Appl1
79	37.8	2.4	13402	3	US-09-105-537-5 Sequence 5, Appl1
80	37.8	2.4	36778	3	US-09-320-878-19 Sequence 19, Appl1
81	37.8	2.4	38506	3	US-09-141-908-1 Sequence 1, Appl1
82	37.8	2.4	38506	4	US-09-657-440-19 Sequence 19, Appl1
83	37.8	2.4	38506	4	US-09-305-640-1 Sequence 1, Appl1
84	37.6	2.4	2650	3	US-10-140-002-119 Sequence 119, App
85	37.6	2.4	2868	4	US-09-774-528-383 Sequence 383, App
86	37.4	2.4	1066	4	US-09-103-840A-1 Sequence 1, Appl1
87	37.2	2.4	648	4	US-09-252-991A-9715 Sequence 9715, Ap
88	37.2	2.4	2150	2	US-08-318-837-1 Sequence 1, Appl1
89	37.2	2.4	3395	3	US-09-428-711A-13 Sequence 13, Appl1
90	37.2	2.4	5561	4	US-09-418-710-28 Sequence 28, Appl1
91	37.2	2.4	5561	4	US-09-839-479-28 Sequence 30, Appl1
92	37.2	2.4	5573	4	US-09-418-710-30 Sequence 30, Appl1
93	37.2	2.4	5573	4	US-08-146-930-1 Sequence 1, Appl1
94	37.2	2.4	6330	2	US-08-458-240-1 Sequence 1, Appl1
95	37.2	2.4	6330	5	PCT-US93-03993-1 Sequence 1, Appl1
96	37.2	2.4	6330	5	US-09-183-861-34 Sequence 34, Appl1
97	37	2.4	516	3	US-09-022-765-34 Sequence 34, Appl1
98	37	2.4	516	4	US-09-551-974A-44 Sequence 44, Appl1
99	37	2.4	516	4	US-09-551-974A-44 Sequence 44, Appl1
100	37	2.4	516	4	US-09-551-974A-44 Sequence 34, Appl1

[illegible]


```

; APPLICANT: Targoff, Ira N.
; APPLICANT: Ge Qun
; TITLE OF INVENTION: Antigens Associated with Polymyositis
; TITLE OF INVENTION: and with Dermatomyositis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Ste. 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: United States
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,295
; FILING DATE: 19920909
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMR# 120C1P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-6555
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1053 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Blood
; IMMEDIATE SOURCE:
; LIBRARY: Human thymocyte lambda gt11
; CLONE: L1
; US-07-945-295-1

Query Match          3.9%; Score 59.8; DB 4; Length 1053;
Best Local Similarity 58.9%; Pred. No. 2e-05;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGAGCGGGGAGCTCATCTGCTGTGACGCGTGCCTC 591
    |||||
DB 248 AGATTACTGTGAGGTGTGCGACAGGAGGGAATTTCTGTGTGACACTTGCCTC 307
    |||||

QY 592 GGGCCTTCACCTGAGCTGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGCTTGA 651
    |||||
DB 308 GTGCTTACCACTCGTCTGCTTGATCTGAGCTTGAACCGGGGTCCAGAGGGGCAATGA 367
    |||||

QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGGTGACGCCCGGGCAGAGGA 706
    |||||
DB 368 GCTGCCCTCACTGTGAGAGGAGGGGCTCAGTGGAGGCCAAGAGGAAGA 422
    |||||

RESULT 6
PCT-US91-06418-2
; Sequence 2, Application PC/TUS9106418
; GENERAL INFORMATION:
; APPLICANT: Oklahoma Medical Research, Foundation, et al
; TITLE OF INVENTION: Antigens Associated with Polymyositis
; TITLE OF INVENTION: and with Dermatomyositis
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
```

```

; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/06418
; FILING DATE: 19910905
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/579023
; FILING DATE: 09-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMR#120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1053 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Sera
; IMMEDIATE SOURCE:
; CLONE: L1
; PCT-US91-06418-2

Query Match          3.9%; Score 59.8; DB 5; Length 1053;
Best Local Similarity 58.9%; Pred. No. 2e-05;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGAGCGGGGAGCTCATCTGCTGTGACGCGTGCCTC 591
    |||||
DB 248 AGATTACTGTGAGGTGTGCGACAGGAGGGAATTTCTGTGTGACACTTGCCTC 307
    |||||

QY 592 GGGCCTTCACCTGAGCTGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGCTTGA 651
    |||||
DB 308 GTGCTTACCACTCGTCTGCTTGATCTGAGCTTGAACCGGGGTCCAGAGGGGCAATGA 367
    |||||

QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGGTGACGCCCGGGCAGAGGA 706
    |||||
DB 368 GCTGCCCTCACTGTGAGAGGAGGGGCTCAGTGGAGGCCAAGAGGAAGA 422
    |||||

RESULT 7
US-08-942-686-1
; Sequence 1, Application US/08942686
; Patent No. 6183988
; GENERAL INFORMATION:
; APPLICANT: Bloch, Donald B.
; TITLE OF INVENTION: LEUKOCYTE-SPECIFIC PROTEIN AND GENE, AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
```


ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,686
FILING DATE: 02-OCT-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/027,347
FILING DATE: 02-OCT-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: GOLDSTEIN, JORGE A.
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 0609.4330001/JAG/BJD
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2905 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 107..2365
US-08-942-686-1

Query Match 3.1%; Score 47.2; DB 3; Length 2905;
Best Local Similarity 64.8%; Pred. No. 0.036;
Matches 70; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 515 CCTGTCCCAAGAAATGAGACAGATGTGCTGCTGTCCGAGCGCGGAGCTCATCTG 574
DB 1816 CCTGTGTATGAGAAACCTGATAGTGTAGAGTGTCCCGGAGCGGAGCTCATCTG 1875

QY 575 CTGTGACGGCTGCGCTCGGAGCTTCCACCTGCGCTGCTGCTCCCTCC 622
DB 1876 TTGCGACACTGTTCAGAGCTTCCATGAGGACTGTCACTCCGCC 1923

RESULT 8
US-09-976-594-1064/c
Sequence 1064; Application US/09976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Furness, Michael
APPLICANT: Buchbinder, Jenny
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
FILE REFERENCE: PA-0041 US
CURRENT APPLICATION NUMBER: US/09/976,594
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240,409
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 1064
LENGTH: 2327
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Inocyte ID No. 6673549 210133.1
US-09-976-594-1064

Query Match 2.9%; Score 44.6; DB 4; Length 2327;
Best Local Similarity 50.7%; Pred. No. 0.15;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 568 TCATCTGCTGTACAGCGCTGCTCCCTCGGAGCTTCCACCTGCGCTGTGCTCCCTCCGCTCC 627
DB 417 TGTGCGCTGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCC 358

QY 628 GGGAGATCCCAAGTGGACCTGAGGTGCTCAGCTGCTGTGAGGCAACAGTCCAGAGAG 687
DB 357 GCGAGGCGAGCGGGGCCCCCGGAGCGGCGGCTCCCAACAGCTTTCGAG 298

QY 688 TCGAGCCCCGGGCAAGAGAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747
DB 297 CGCTGCTGGGAGAGCGGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCT 238

QY 748 CGAGGCTTAGTGTGCGCGGAGAGAGAGTAA 778
DB 237 TGTGCTTCATCAGAGATAGAGAGACTAG 207

RESULT 9
US-08-340-203A-1/c
Sequence 1; Application US/08340203A
Patent No. 5756668
GENERAL INFORMATION:
APPLICANT: Baylin, Stephen B.
TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,203A
FILING DATE: 15-NOV-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: 07265/039001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4616 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: HIC-1 polynucleotide
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4616
US-08-340-203A-1

Query Match 2.9%; Score 44.6; DB 1; Length 4616;
Best Local Similarity 50.7%; Pred. No. 0.19;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 568 TCATCTGCTGTACAGCGCTGCTCCCTCGGAGCTTCCACCTGCGCTGTGCTCCCTCCGCTCC 627
DB 2211 TGTGCGCTGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCCCGCTCC 2152

QY 628 GGGAGATCCCAAGTGGAGAGCTGTGCTCAGCTGCTGTGAGGCAACAGTCCAGAGAG 687

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

RESULT 12

US-09-085-407-1/c
Sequence 1, Application US/09085407

Patent No. 6103877

GENERAL INFORMATION:

APPLICANT: Baylin, Stephen B.

APPLICANT: Males, Michele M.

TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: California

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/085,407

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/340,203

FILING DATE: 15-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Hallie, Ph.D., Lisa A.

REGISTRATION NUMBER: P-38,347

REFERENCE/DOCKET NUMBER: 07265/039001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 455-5100

TELEFAX: (619) 455-5110

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4616 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

IMMEDIATE SOURCE:

CLONE: HIC-1 polynucleotide

FEATURE:

NAME/KEY: CDS

LOCATION: 1..4616

US-09-085-407-1

Query Match

Best Local Similarity 2.9%; Score 44.6; DB 3; Length 4616;

Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.9%; Score 44.6; DB 3; Length 4616;

Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.9%; Score 44.6; DB 3; Length 4616;

Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

RESULT 13

US-09-920-923B-36

Sequence 36, Application US/09920923B

Patent No. 6677134

GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Teyanov, Yuri

TITLE OF INVENTION: Fermentative Carotenoid Production

FILE REFERENCE: 15464 US (C38435/125944)

CURRENT APPLICATION NUMBER: US/09/920,923B

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 08/980,832

PRIOR FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: Patentin version 3.1

SEQ ID NO 36

LENGTH: 882

TYPE: DNA

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: E-396

US-09-920-923B-36

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001

Query Match

Best Local Similarity 2.8%; Score 42.6; DB 4; Length 882;

Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Db 2031 TCCTGCTTCATCCAGCATAGAGAGACTAG 2001


```

; TITLE OF INVENTION: A new method of identifying non-host plant disease
; TITLE OF INVENTION: resistance genes
; FILE REFERENCE: r gene patent
; CURRENT APPLICATION NUMBER: US/09/387,286
; EARLIER FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: 60/098,402
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 3612
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
; US-09-387-286-60

Query Match
Best Local Similarity 66.7%; Score 42; DB 4; Length 3612;
Matches 60; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 521 CCAGAGATGAGAGAGAGTGTGCGGAGCGGCGGAGCTCATCTGTGTGA 580
DB 1961 CATAACAGTGAATGATGTGATGATGATGCGGAGTGGGAGAACTGATTGCTGTGA 2020
QY 581 CGGCTGCGCTCGGCGCTTCACCTGCGCTG 610
DB 2021 AGGGTGTCTCGGCGCTTTCATGACAGCTTG 2050

RESULT 21
PCT-US91-07635-5
; Sequence 5, Application PC/TUS9107635
; GENERAL INFORMATION:
; APPLICANT: OPPERMANN, HERMANN
; APPLICANT: OZKAYNAK, ERGIN
; APPLICANT: RUBER, DAVID C
; APPLICANT: KUBERASAMPATH, THANGAVEL
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HUMWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/07635
; FILING DATE: 19911018
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDWARD R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-056PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1941 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HOMO SAPIENS
; TISSUE TYPE: HIPPOCAMPUS
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 507..1703
; OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product= "HOP2"
; OTHER INFORMATION:
; PCT-US91-07635-5

Query Match
Best Local Similarity 51.6%; Score 41.6; DB 5; Length 1941;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 566 GCTCATCTGCTGTGACGAGCTGCGCTTCGACCTGCGCTGCTGCTCCCTGCT 625
DB 533 GCTCTCGGCTCTGCGGCTATGCGCGTGTGGCGGCGGCGGCGGCTGCGACCCCGCC 592
QY 626 CCGGAGATCCCAAGTGGAGCTGAGAGTGTTCAGCTGCTGACGAGCAACAGTCCAGA 685
DB 593 CGGCTGTCCCGACGACGCTGTGGGCGCGCGACCGGAGCGTGAAGCGGAGATCTGGC 652
QY 686 GGTGACAGCCCGGCGGAGAGAGCCCGGCGCCGAGAGCCCGTGGAGACCCCGCTCC 745
DB 653 GGTGCTCGGCTGTGCTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 712
QY 746 CCCG 749
DB 713 CGCG 716
```

```

RESULT 22
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMWU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```



```
GENERAL INFORMATION:
APPLICANT: KIKLY, KRISTINE K
APPLICANT: GROSS, MITCHELL S
APPLICANT: HURLE, MARK R
TITLE OF INVENTION: HUMAN REQUIEM
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: RAYNER & PRESTITA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/881,857
FILING DATE: 24-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/021,229
FILING DATE: 26-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: PRESTITA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: ATG-50013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1556 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-881-857-1

Query Match
Best Local Similarity 59.0%; Score 40.2; DB 2; Length 1556;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 554 GGACGGCGGGAGCTCATCTGTGTGACGCGCTGCGGCGCTTCCACCTGCGCTGCT 613
DB 1073 GAATGACGACCACTGCTCTTCTGTGATGATGATGATGATGATGATGATGATGAT 1132

QY 614 GTCCCTCCGCTCCGGAGATCCCAAGTGGACCTGGAGGTGCTCCAGCTGCTGCA 670
DB 1133 CACCCGCTCATGTGTGACGCCCTGAGAGAGTTGAGACTGCGCACCTGTGTCTGA 1189

RESULT 29
US-09-233-342A-1
Sequence 1, Application US/09233342A
Patent No. 6207803
GENERAL INFORMATION:
APPLICANT: KIKLY, KRISTINE K.
APPLICANT: GROSS, MITCHELL S.
APPLICANT: HURLE, MARK ROBERT
TITLE OF INVENTION: HUMAN REQUIEM
FILE REFERENCE: ATG-50013-1
CURRENT APPLICATION NUMBER: US/09/233,342A
CURRENT FILING DATE: 1999-01-19
PRIOR APPLICATION NUMBER: 08/881,857
PRIOR FILING DATE: 1997-06-24
PRIOR APPLICATION NUMBER: 60/021,299
PRIOR FILING DATE: 1996-06-26
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Fastseq for Windows Version 3.0
```

```
SEQ ID NO 1
LENGTH: 1556
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-233-342A-1

Query Match
Best Local Similarity 59.0%; Score 40.2; DB 3; Length 1556;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 554 GGACGGCGGGAGCTCATCTGTGTGACGCGCTGCGGCGCTTCCACCTGCGCTGCT 613
DB 1073 GAATGACGACCACTGCTCTTCTGTGATGATGATGATGATGATGATGATGATGAT 1132

QY 614 GTCCCTCCGCTCCGGAGATCCCAAGTGGACCTGGAGGTGCTCCAGCTGCTGCA 670
DB 1133 CACCCGCTCATGTGTGACGCCCTGAGAGAGTTGAGACTGCGCACCTGTGTCTGA 1189

RESULT 30
US-09-620-312D-389/c
Sequence 389, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunging
APPLICANT: Wang, Dairui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE REFERENCE: 784C1P2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
SOFTWARE: pc_FL_genes Version 1.0
NUMBER OF SEQ ID NOS: 1105
SEQ ID NO 389
LENGTH: 1682
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (25) .. (1197)
US-09-620-312D-389

Query Match
Best Local Similarity 49.3%; Score 40.2; DB 4; Length 1682;
Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 964 GTACGAGCGTGTGCGGTACTGACGTGCGCGCTTCCACCTGCGCGCTGCTGCT 1023
DB 217 GCACGAGACACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 158

QY 1024 CAGCGGACCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1083
DB 157 CCCGCTCATGTGTGACGCCCTGAGAGAGTTGAGACTGCGCACCTGTGTCTGA 98
```

Qy 1084 CCCGAGCCCTGTGAGAGGGGAGTGTGAGCCCGCCCGCCCTGAGCCCTGAGGCTG 1143
Db 97 AGCTGAGGCGCCGCGCGCGGTAGCTGCGCCCAAGCCCTGACCACTACCGCTGCGCA 38
Qy 1144 CCAAGATGACACTGCGCACTGACGAGCCCGCTC 1176
Db 37 CCCGAGCGCTCATGTGTGAGACACACTGCGCGCTC 5

RESULT 31
US-08-969-106-9
; Sequence 9, Application US/08969106
; Patent No. 5986055
; GENERAL INFORMATION:
; APPLICANT: Yang, M.
; APPLICANT: Nandabalan, K.
; APPLICANT: Schulz, V.
; TITLE OF INVENTION: CDK2 INTERACTIONS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Penlie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/969,106
; FILING DATE: 13-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7934-057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-969-106-9

Query Match 2.6%; Score 40.2; DB 2; Length 2399;
Best Local Similarity 59.0%; Pred. No. 1.8;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
Qy 554 GGACGGCGGGAGCTCATGTGAGCGGCTGCGGCGCTTCCAGCTGAGCGCTG 613
Db 1016 GAATGACGACGAGTCTCTTCTGTGATGACTGCGATCGTGGCTACCACTGACTGTCT 1075
Qy 614 GTCCCTCCGCTCCGGGAGATCCCAAGTGGAGCTGAGGTGCTCCAGCTGCTGCA 670
Db 1076 CACCCCGTCCATGTGTGAGCGCCCTGAAGAGATTGAGACTGCACTGTGTCTGA 1132

RESULT 32
US-09-338-125-9
; Sequence 9, Application US/09338125
; Patent No. 6521412
; GENERAL INFORMATION:
; APPLICANT: Yang, M.
; APPLICANT: Nandabalan, K.
; APPLICANT: Schulz, V.

; TITLE OF INVENTION: CDK2 INTERACTIONS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Penlie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/338,125
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/969,106
; FILING DATE: 13-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7934-057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-338-125-9

Query Match 2.6%; Score 40.2; DB 4; Length 2399;
Best Local Similarity 59.0%; Pred. No. 1.8;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
Qy 554 GGACGGCGGGAGCTCATGTGAGCGGCTGCGGCGCTTCCAGCTGAGCGCTG 613
Db 1016 GAATGACGACGAGTCTCTTCTGTGATGACTGCGATCGTGGCTACCACTGACTGTCT 1075
Qy 614 GTCCCTCCGCTCCGGGAGATCCCAAGTGGAGCTGAGGTGCTCCAGCTGCTGCA 670
Db 1076 CACCCCGTCCATGTGTGAGCGCCCTGAAGAGATTGAGACTGCACTGTGTCTGA 1132

RESULT 33
US-09-313-294A-604
; Sequence 604, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 604
; LENGTH: 293
; TYPE: DNA
; ORGANISM: Zea mays
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700549516H1
US-09-313-294A-604

Query Match 2.6%; Score 40; DB 4; Length 293;
Best Local Similarity 60.6%; Pred. No. 1;
Matches 86; Conservative 0; Mismatches 50; Indels 6; Gaps 1;

QY 526 AGAATGAGACGAGTGTGCGGAGCGGAGGAGCTCATCTGTGAGCGCT 585
DB 18 AAAAGATATACCTGTGATCATGTGCGATGTGTAACTGATATCTGACAAATT 77
QY 586 GCCCTGGGCTTCCACCTGAGCTGCTGCTCCCTCCGTCGCGAGATCCCAAGTGGGA 645
DB 78 GTCTGCTAGTATACCATCAAGCTGTGCTTGTG-----CCAGATATTCAGATGGA 131
QY 646 CCTGAGAGTCTCCAGCTGCTT 667
DB 132 ACTGTACTCTCTAGCTGCTT 153

RESULT 34
US-09-252-991A-11986
; Sequence 11986, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11986
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11986

Query Match 2.6%; Score 40; DB 4; Length 762;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 97; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1096 TGAAGGGGCTGCTGAGCCCGCCGCTGAGCCCTGAGCCCTGCAAGATGACA 1155
DB 386 TGGTCGAGGTGCTGATGCGCGAAGAACATTCGCGCCGCTCCTACGAAGTGA 445
QY 1156 CTGCAAGTACAGAGCCGCTGTCACAGGATGACTGATGCTTCTGAGCGACACA 1215
DB 446 CACCGAAGGTGAAGCGGCTGCGCGCCGCGAGGGAGTCTATCCACGCGGCACA 505
QY 1216 CTTGATGAGTCTGCTGAGTGGGCGCATCGAGCATGAGCCGCTGCGGCGCCCTTCC 1275
DB 506 CCATGAGACGACTTCAAGGCTGCTGTGATGATGCTGTGCTGCGGACATCAGATG 565
QY 1276 CCTCTGACCCC 1287
DB 566 CGTCTTCGCCC 577

RESULT 35
US-09-252-991A-11956
; Sequence 11956, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11956
; LENGTH: 1125
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11956

Query Match 2.6%; Score 40; DB 4; Length 1125;
Best Local Similarity 50.5%; Pred. No. 1.6;
Matches 97; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1096 TGAAGGGGCTGCTGAGCCCGCCGCTGAGCCCTGAGCCCTGCAAGATGACA 1155
DB 576 TGGTCGAGGTGCTGATGCGCGAAGAACATTCGCGCCGCTCCTACGAAGTGA 635
QY 1156 CTGCAAGTACAGAGCCGCTGTCACAGGATGACTGATGCTTCTGAGCGACACA 1215
DB 636 CACCGAAGGTGAAGCGGCTGCTGCGCGCCGCGAGGGAGTCTATCCACGCGGCACA 695
QY 1216 CTTGATGAGTCTGCTGAGTGGGCGCATCGAGCATGAGCCGCTGCGGCGCCCTTCC 1275
DB 696 CCATGAGACGACTTCAAGGCTGCTGTGATGATGCTGTGCTGCGGACATCAGATG 755
QY 1276 CCTCTGACCCC 1287
DB 756 CGTCTTCGCCC 767

RESULT 36
US-08-680-506-5
; Sequence 5, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 5
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-5

Query Match 2.6%; Score 39.6; DB 3; Length 954;
Best Local Similarity 44.9%; Pred. No. 1.9;
Matches 150; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 985 CTCAGTGGCGCGCTGCTTCACTGCGCTGCACTTCCAGCGACCTTCCGAGCCG 1044
DB 5 CGCCGCTGCGCGCTGCTGCTGCTGCGCTTCTGTGCGGTGCGAGCGCGCGGCTGG 64
QY 1045 GGAAGGGGCTGCGTGCAGATCTGCTGAGAGAGTACCCACGCCCTGTGAGAGGGG 1104
DB 65 GCGGCTACTTGGGACACAAAGTCCGCTACGAGAGGTAAACCGACCTGCGGAGAGACC 124
QY 1105 TGTCTGCCCCCAGCCCCCGCGCTGAGCCCTGAGCGCTGCAAGATGACATGCGAGTC 1164
DB 125 CGCTCTCCCTGGGCGGACAGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCTGTC 184
QY 1165 ACGAGCCGCTGTGACAGGATGACTGAGTCCCTTCTGAGCGACACACTTTCGATG 1224
DB 185 AGCTCCGCGCGCTCTCCGCGCACCGGACCGGCTACCCACGCGCGGAGAAATCCCGCGC 244
QY 1225 GCATCTGAGTGGGCGATTCAGAGCATGAGCGCGCGCGCGCGCGCTTCCCTCTGAC 1284

Db 325 TGCCGAGCTGCACGGCCGCTCCGCGCGCCGCGCCGCTCTGCTCCGCGCGCGCG 384
QY 1285 CCCAGATGGCCGGGACATGCAGCTTGATGAGAG 1318
Db 385 CGCTGGCCGCTGGCCGATGTGTACGAGAGAG 418

RESULT 40
US-08-785-310A-4/c
Sequence 4, Application US/08785310A
Patent No. 5840532
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L.
APPLICANT: Russell, David W.
TITLE OF INVENTION: Neuronal PAS Domain Protein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,310A
FILING DATE: 21-JAN-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UTSD:1226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4184 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-785-310A-4

Query Match 2.6%; Score 39.6; DB 2; Length 4184;
Best Local Similarity 50.4%; Pred. No. 3.1;
Matches 122; Conservative 0; Mismatches 119; Indels 1; Gaps 1;
QY 276 CAGCTCGAAGAAACCGGGTTTCTCCCAATAGGATGCGCCCGGGGGGTGCTGTGG 335
Db 251 CACTCTCCAGGGCTCCCGGCTCCCTACGTTGGGATGCTCGGTGGGCTCTTGAAG 192
QY 336 AGACGATGATGGGAAACAGT-GGTGAGGGCAGAAATTCAGGCTTCGACGATGG 394
Db 191 GCGAGGGGCAATCTTGAAGCTTCTGTCGGGCGACGATACCCCTTAGACCTCCG 132
QY 395 AGCAGGGCAGAGATCGGAGTTCAAGTACCCAGAAATGCTGTGGGGAGCTGTTTGG 454
Db 131 AGGAGCCCGGAATCCGATGTTTGGGAGCTTAGTCTCTTGGCAGCAGCTCCG 72
QY 455 GAAGAGGTGCTCTCAGAGAGGTGCTGACCCGACCCAGTGTGATGGGCTCTCTTG 514
Db 71 GAGCCACCGCTGTCTGTGCTGAGCGGTTGGAGAGACCCGAGCCAGGGCGGCTG 12
QY 515 CC 516
Db 11 CC 10

RESULT 41
US-08-680-506-2
Sequence 2, Application US/08680506C
Patent No. 6008013
GENERAL INFORMATION:
APPLICANT: Reynolds, Paul R.
TITLE OF INVENTION: CHONDROCYTE PROTEINS
FILE REFERENCE: 176/60091
CURRENT APPLICATION NUMBER: US/08/680,506C
CURRENT FILING DATE: 1996-07-08
EARLIER APPLICATION NUMBER: 60/021,672
EARLIER FILING DATE: 1996-07-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
TYPE: DNA
LENGTH: 5027
ORGANISM: Gallus gallus
US-08-680-506-2

Query Match 2.6%; Score 39.6; DB 3; Length 5027;
Best Local Similarity 44.9%; Pred. No. 3.3;
Matches 150; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 985 CTCACCTGCGCCGCTGCTTCCTCACTGCGGCTGCCACTTCCAGCCGACCTCCGCGCG 1044
Db 85 CCGCGTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 144
QY 1045 GAGCGGCGCTGCGCTGCGACATCTGCTCAGAGAGTGACCCGAGCCCTGTGAGAGGG 1104
Db 145 GCGGCTACTTGGGACACAGTCCGCTTACGAGAGAGTGAACCGACCTGGGAGAGACC 204
QY 1105 TCGTGGCCCCAGCCCGCCGCTGCGCCCTGAGGCTGCGCAAGATGCACTGCCAGTC 1164
Db 205 CGGTGCTCCGCGGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 264
QY 1165 AGGAGCGGCTGTGACAGAGGATGACCTGAGATCCCTTTGAGCGAGCACACTTGATG 1224
Db 265 ACTCCGCGCGCTGCTGCTGCGCACGAGCACCGGCTACCCAGCGCGGCAAAATCCGCGCC 324
QY 1225 GCATCTGAGTGGGACATCCAGAGATGAGCCGCTGCGGCGCGCGCGCGCGCGCGCTG 1284
Db 325 TGCCGAGCTGCACGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384
QY 1285 CCCAGATGCGCGGACATGACAGCTCTGATGAGAG 1318
Db 385 CGCTGGCCGCTGGCCGATGTGTACGAGAGAGAG 418

RESULT 42
US-09-679-279-1/c
Sequence 1, Application US/09679279
Patent No. 6524841
GENERAL INFORMATION:
APPLICANT: McDaniel, Robert
APPLICANT: Volchegursky, Yanina
TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
FILE REFERENCE: 300623004700
CURRENT APPLICATION NUMBER: US/09/679,279
CURRENT FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/158,305
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 60/190,024
PRIOR FILING DATE: 2000-03-17
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 47981
TYPE: DNA
ORGANISM: Micromonospora megalomicea
FEATURE:

```
? NAME/KEY: CDS
? LOCATION: (1)....(144)
? OTHER INFORMATION: megBVI[megT], TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
? OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (328)...(2061)
? OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,
? OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
? OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (2072)...(3382)
? OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),
? OTHER INFORMATION: TDP-megosamine glycosyltransferase;
? OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (3462)...(4634)
? OTHER INFORMATION: megG[megY], mycarosyl acyltransferase, mycarose O-acyltransferase
? OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (4651)...(5775)
? OTHER INFORMATION: megDII, deoxysugar transferase (eryCI, DnrJ homolog),
? OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;
? OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (5822)...(6595)
? OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
? OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (6592)...(7197)
? OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmuv hc
? OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
? OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (7220)...(8206)
? OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dmuv homolog),
? OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
? OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (8228)...(9220)
? OTHER INFORMATION: megBII-1[megDVII], TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
? OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (9226)...(10479)
? OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
? OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (10483)...(11424)
? OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
? OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
? OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
? NAME/KEY: CDS
? LOCATION: (12181)...(22821)
? OTHER INFORMATION: megAI, SEQ ID NO: 13= translated amino acid sequence
? NAME/KEY: misc_feature
? LOCATION: (12505)...(13470)
? OTHER INFORMATION: megAI, AT-L
? NAME/KEY: misc_feature
? LOCATION: (13576)...(13791)
? OTHER INFORMATION: megAI, ACP-L
? NAME/KEY: misc_feature
? LOCATION: (13849)...(15126)
? OTHER INFORMATION: megAI, KS1
? NAME/KEY: misc_feature
? LOCATION: (15427)...(16476)
? OTHER INFORMATION: megAI, ATI
? NAME/KEY: misc_feature
? LOCATION: (17155)...(17694)
? OTHER INFORMATION: megAI, KRI
? NAME/KEY: misc_feature
? LOCATION: (17947)...(18207)
? OTHER INFORMATION: megAI, ACP1
? NAME/KEY: misc_feature
? LOCATION: (18268)...(19548)

? OTHER INFORMATION: megAI, KS2
? NAME/KEY: misc_feature
? LOCATION: (19876)...(20910)
? OTHER INFORMATION: megAI, AT2
? NAME/KEY: misc_feature
? LOCATION: (21517)...(22053)
? OTHER INFORMATION: megAI, KR2
? NAME/KEY: misc_feature
? LOCATION: (22318)...(22575)
? OTHER INFORMATION: megAI, ACP2
? NAME/KEY: CDS
? LOCATION: (22867)...(33555)
? OTHER INFORMATION: megAII; SEQ ID NO: 14= translated amino acid sequence
? NAME/KEY: misc_feature
? LOCATION: (32957)...(24237)
? OTHER INFORMATION: megAII, KS3
? NAME/KEY: misc_feature
? LOCATION: (24544)...(25581)
? OTHER INFORMATION: megAII, AT3
? NAME/KEY: misc_feature
? LOCATION: (26230)...(26733)
? OTHER INFORMATION: megAII, KR3 (inactive)
? NAME/KEY: misc_feature
? LOCATION: (26998)...(27258)
? OTHER INFORMATION: megAII, ACP3
? NAME/KEY: misc_feature
? LOCATION: (27393)...(28590)
? OTHER INFORMATION: megAII, KS4
? NAME/KEY: misc_feature
? LOCATION: (28897)...(29931)
? OTHER INFORMATION: megAII, AT4
? NAME/KEY: misc_feature
? LOCATION: (29953)...(30477)
? OTHER INFORMATION: megAII, DH4
? NAME/KEY: misc_feature
? LOCATION: (31396)...(32244)
? OTHER INFORMATION: megAII, BR4
? NAME/KEY: misc_feature
? LOCATION: (32257)...(32799)
? OTHER INFORMATION: megAII, KR4
? NAME/KEY: misc_feature
? LOCATION: (33052)...(33312)
? OTHER INFORMATION: megAII, ACP4
? NAME/KEY: CDS
? LOCATION: (33666)...(43271)
? OTHER INFORMATION: megAII; SEQ ID NO: 15= translated amino acid sequence
? NAME/KEY: misc_feature
? LOCATION: (33780)...(35027)
? OTHER INFORMATION: megAII, KS5
? NAME/KEY: misc_feature
? LOCATION: (35385)...(36419)
? OTHER INFORMATION: megAII, AT5
? NAME/KEY: misc_feature
? LOCATION: (37068)...(37604)
? OTHER INFORMATION: megAII, KR5
? NAME/KEY: misc_feature
? LOCATION: (37860)...(38120)
? OTHER INFORMATION: megAII, ACP5
? NAME/KEY: misc_feature
? LOCATION: (38187)...(39470)
? OTHER INFORMATION: megAII, KS6
? NAME/KEY: misc_feature
? LOCATION: (39795)...(40811)
? OTHER INFORMATION: megAII, AT6
? NAME/KEY: misc_feature
? LOCATION: (41406)...(41936)
? OTHER INFORMATION: megAII, KR6
? NAME/KEY: misc_feature
? LOCATION: (42168)...(42425)
? OTHER INFORMATION: megAII, ACP6
? NAME/KEY: misc_feature
? LOCATION: (42585)...(43271)
? OTHER INFORMATION: megAII, TE
```


;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 13605
;; LENGTH: 804
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13605

Query Match
Best Local Similarity 47.2%; Score 38.8; DB 4; Length 804;
Matches 118; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1043 CCGGACGGGCGCTGCTGATCTCTCAGAGACGTGACCCCGCTGTGAGG 1102
DB 303 CCGGCTGGGGCGGACAGCATCTGCGCTGCGCACTTCCGCGCGGATGCGGCC 362
QY 1103 GGTGCTGGCCCCAGGCCCCCGCTGCGCTGCGCTGCGCAAGATGACATGCCAG 1162
DB 363 AGGTGGGACATGAGCGGATGACACGCGCGCGGATTCGCGTCTATCCGTGCGCCAG 422
QY 1163 TCACGAGCGCGCTCTGACAGGAGTGAAGTCCCTTCTGAGGAGACACCTTCA 1222
DB 423 TCACAAAGGCGGCACTGCGCGCTGCTCATGAGTGTGCTTTCAGAGGCGCTCCG 482
QY 1223 TGGCATCTGCAATGAGGCGCATTCAGAGATGCGCGCGCGCGCCCTTCCCTCTG 1282
DB 483 CCGCGAGCGGACGCGCGCGGAGAGCGCGCTGCGCGGTTGAGAAAGCGGCTCGGCGA 542
QY 1283 ACCCCGATG 1292
DB 543 GCGCTGCTG 552

RESULT 56
US-09-252-991A-13494
; Sequence 13494, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13494
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13494

Query Match
Best Local Similarity 47.2%; Score 38.8; DB 4; Length 984;
Matches 118; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1043 CCGGACGGGCGCTGCTGATCTCTCAGAGACGTGACCCCGCTGTGAGG 1102
DB 145 CCGGCTGGGGCGGACAGCATCTGCGCTGCGCACTTCCGCGCGGATGCGGCC 204
QY 1103 GGTGCTGGCCCCAGGCCCCCGCTGCGCTGCGCTGCGCAAGATGACATGCCAG 1162
DB 205 AGGTGGGACATGAGCGGATGACACGCGCGCGGATTCGCGTCTATCCGTGCGCCAG 264
QY 1163 TCACGAGCGCGCTCTGACAGGAGTGAAGTCCCTTCTGAGGAGACACCTTCA 1222
DB 265 TCACAAAGGCGGCACTGCGCGCTGCTCATGAGTGTGCTTTCAGAGGCGCTCCG 324

QY 1223 TGGCATCTGCAATGAGGCGCATTCAGAGATGCGCGCGCGCGCCCTTCCCTCTG 1282
DB 325 CCGGAGCGGACAGGCGCGGAGCGCGCTGCGCGGTTTGAGAAAGCGGCTTCGCGCA 384
QY 1283 ACCCCGATG 1292
DB 385 GCGCTGCTG 394

RESULT 57
US-09-252-991A-4137/C
; Sequence 4137, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4137
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4137

Query Match
Best Local Similarity 44.2%; Score 38.8; DB 4; Length 1104;
Matches 160; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 360 GGTCAAGGCGAATTTCAAGGCGCTGCGAGCATGAGGAGGAGGAGAGACTGGAGTTCA 419
DB 522 GACATGCGGATCATCTCTCTGCTGCGCGCGCGCGCTGAGATGATGACATGCTGATG 463
QY 420 GGTACCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 479
DB 462 AGTGTGCGGAGAGCGCGCGCGGAGAAATCGCGCTGCGCATGCGCTGCGCGCGCGCG 403
QY 480 CTGACCCCAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 539
DB 402 GACATCGGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 343
QY 540 TGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599
DB 342 GCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 283
QY 600 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
DB 282 TCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 223
QY 660 ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 719
DB 222 TTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 163
QY 720 GA 721
DB 162 GA 161

RESULT 58
US-09-252-991A-4071
; Sequence 4071, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136


```

; APPLICANT: Huang, Huang
; TITLE OF INVENTION: Generation of Human Cytomegalovirus Yeast Artificial Chromosome
; FILE REFERENCE: 98,299
; CURRENT APPLICATION NUMBER: US/09/765,400
; CURRENT FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 229354
; TYPE: DNA
; ORGANISM: Human cytomegalovirus;
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Human cytomegalovirus strain AD169 (GenBank X17403.1)
US-09-765-400-64
```

```

Query Match          2.5%; Score 38.6; DB 4; Length 229354;
Best Local Similarity 49.8%; Pred. No. 21;
Matches 124; Conservative 0; Mismatches 124; Indels 1; Gaps 1;
```

```

QY 946 GCGGGGTGTGCGGAGATGTGATCGAGCTGCTGCGGTGTACTCACTGCGCGCTGCTTCC 1005
    |||
DB 93012 GCGGAGTCCGCGGACAGGATTCGCGCGGTGCTGCTGCGGACGCTGCGCGCTCCGCG 93071
    |||
QY 1006 ACTGCGGCTGCCACTTCCAGCCGCGACCTCCGCGCGGAGCGGCGCTGCGTGAAT 1065
    |||
DB 93072 TGCCTCCCGCCCTACCCCGCCACCTCCCGCGGAGCGGCGCGGCTGCTCCGCG 93131
    |||
QY 1066 CCGTCTCAGAGAGCGTACGCCCGCTGTGAGGGGGTGTGCGCGCGCGCGCGCG 1125
    |||
DB 93132 CCGTCTCCACCGCGCTG-GAGCAGCATCCGCGCGCTGCGCGCGCGCGCGCGCG 93190
    |||
QY 1126 GCCTGCGCGCGCTGCGCGCTGCGCAGAGATGACATGACGAGCGCGCTGCGACAGG 1185
    |||
DB 93191 GCTCCGAGACTCGCGCGGCGGCTCCCTCCCGCGCGCTGCGAGCGCGCGCGCGCG 93250
    |||
QY 1186 ATGACCTGG 1194
    |||
DB 93251 CCGGCGCGG 93259
    |||
```

```

RESULT 62
US-09-705-400-64
; Sequence 64, Application US/09705400
; Patent No. 6692954
; GENERAL INFORMATION:
; APPLICANT: Ghazal, Peter
; APPLICANT: Huang, Huang
; TITLE OF INVENTION: Generation of Human Cytomegalovirus Yeast Artificial Chromosome
; FILE REFERENCE: 98,299
; CURRENT APPLICATION NUMBER: US/09/705,400
; CURRENT FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 229354
; TYPE: DNA
; ORGANISM: Human cytomegalovirus
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Human cytomegalovirus strain AD169 (GenBank X17403.1)
US-09-705-400-64
```

```

Query Match          2.5%; Score 38.6; DB 4; Length 229354;
Best Local Similarity 49.8%; Pred. No. 21;
Matches 124; Conservative 0; Mismatches 124; Indels 1; Gaps 1;
```

```

QY 946 GCGGGGTGTGCGGAGATGTGATCGAGCTGCTGCGGTGTACTCACTGCGCGCTGCTTCC 1005
    |||
DB 93012 GCGGAGTCCGCGGACAGGATTCGCGCGGTGCTGCTGCGGACGCTGCGCGCTCCGCG 93071
    |||
```

```

QY 1006 ACTGCGGCTGCCACTTCCAGCGCGGACCTCCCGCGCGGAGCGGCGCTGCGTGAAT 1065
    |||
DB 93072 TGCCTCCCGCCCTACCCCGCCACCTCCCGCGGAGCGCGCGCGGCTGCTCCGCG 93131
    |||
QY 1066 CCGTCTCAGAGAGCGTACGCCCGCTGTGAGGGGGTGTGCGCGCGCGCGCGCG 1125
    |||
DB 93132 CCGTCTCCACCGCGCTG-GAGCAGCATCCGCGCGCTGCGCGCGCGCGCGCGCG 93190
    |||
QY 1126 GCCTGCGCGCGCTGCGCGCTGCGCAGAGATGACATGACGAGTCAAGCGCGCTGCGACAGG 1185
    |||
DB 93191 GCTCCGAGACTCGCGCGGCGGCTCCCTCCCGCGCGCTGCGAGCGCGCGCGCGCG 93250
    |||
QY 1186 ATGACCTGG 1194
    |||
DB 93251 CCGGCGCGG 93259
    |||
```

```

RESULT 63
US-09-724-797-21
; Sequence 21, Application US/09724797
; Patent No. 673398
; GENERAL INFORMATION:
; APPLICANT: Jon S. THORSON
; TITLE OF INVENTION: MICROMOSPORE ECHINOSPORE GENES
; TITLE OF INVENTION: ENCODING FOR BIOSYNTHESIS OF
; FILE REFERENCE: 2653-40
; CURRENT APPLICATION NUMBER: US/09/724,797
; CURRENT FILING DATE: 2000-11-28
; PRIOR FILING DATE: 1998-12-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Bacteria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1683)
US-09-724-797-21
```

```

Query Match          2.5%; Score 38.4; DB 4; Length 1683;
Best Local Similarity 48.6%; Pred. No. 4.5;
Matches 105; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
```

```

QY 555 GACGGGGGAGTATCTGTGACGAGCTGCGCGCGCTTCCAGCTGAGCTGCGCTG 614
    |||
DB 41 GCGTGGGGGACACCTGCGGCTGCTGCGCGCGCGCGCGCTCCACAGGCGGACCGCGCG 100
    |||
QY 615 TCCCTTCCGCTCCGAGATGCCCAAGTGGAGCTGAGAGTGTCTGAGCTGCGAGGA 674
    |||
DB 101 CCGCTGTGCGCGCGCGGAGCGCGAGTGTGCTGCGGCTGAGCAGCGGCGAGTATGAGCT 160
    |||
QY 675 ACAATTCAGAGATGACAGCGCGCGGAGAGAGCGCGCGCGCGCGCGAGAGCGAGG 734
    |||
DB 161 ACCGCGCGGAGACCGACCGCGGAGCTGAGCGCGCGCTGCGCGCGCTGCGAGCTGAGCA 220
    |||
QY 735 ACCCGCTCCCGCGCGGCTTGTGAGTGTGCGCGGAGAG 770
    |||
DB 221 CCGAGTCCCTTCGACATCAACGCGCGGCTGCGGCG 256
    |||
```

```

RESULT 64
US-09-077-940A-3
; Sequence 3, Application US/09077940A
; Patent No. 6576441
; GENERAL INFORMATION:
; APPLICANT: KIMURA, Toru et al.
; TITLE OF INVENTION: NOVEL SEMAPHORIN 2 AND GENE ENCODING THE SAME
; FILE REFERENCE: 0020-4426P
; CURRENT APPLICATION NUMBER: US/09/077,940A
; CURRENT FILING DATE: 1998-06-05
```

NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patentin version 3.1
SEQ ID NO 3
LENGTH: 3524
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: (1) .. (38)
OTHER INFORMATION:
NAME/KEY: 3'UTR
LOCATION: (2706) .. (3524)
OTHER INFORMATION:
NAME/KEY: CDS
LOCATION: (39) .. (2702)
OTHER INFORMATION:
US-09-077-940A-3

Query Match 2.5%; Score 38.4; DB 4; Length 3524;
Best Local Similarity 47.0%; Pred. No. 5.8;
Matches 154; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 423 ACCCAGAGTCTGCTGGGGAGCTGTTTGGAGAGAGTGCTCTCAGAGAGGTCTG 482
DB 1887 ACCGTGGGCTGTGTTCTGTGGGCTCTCCGTGAGCGCGGAGCTGCGCCGGCGCAGGACAG 1946
QY 483 CACCCAGCC---CAGTGTGATGGGCGTCTTGTGCGCAGAGATGAGAGAGAG 539
DB 1947 GAGGCGATCTGGCGGAGCGGGGCGGAGCGGCTGAGCTCAGCCGCTGGGCGAG 2006
QY 540 TGTGCGGTGTGTCGGGACGGCGGAGAGCTATCTGTGAGCGGCTGCGGCGCTTC 599
DB 2007 CCGAGGGCGCAGAGGTCTCCGGGGGCGGGGCGGAGCGGTGCGGTGCGCGCGGCGTTCC 2066
QY 600 CACTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGACTGGAGTCTCC 659
DB 2067 CCGAGGCTCTGTGCGGCGCTCTGATGAGAAAGGCTGGGCGCAAGGCGCAGCTGTCAG 2126
QY 660 AGCTGCTGCGAGCAAGTCCAGAGAGGAGCGCCCGGCGAGAGCGCCGCGCCAG 719
DB 2127 GCGGGGCGCCAGAGCTGAGCTGGGCGTGGCCAGCCCGAGAGAGCGCGCTGCGG 2186
QY 720 GAGCCACCCGCTGAGAGCCCGCTCCCGC 747
DB 2187 CAGAGCGCTGCGCCAGCTCCGAGCCCGC 2214

RESULT 65

US-10-140-002-543
Sequence 543, Application US/10140002
Patent No. 6725730
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Guirney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3350R1C59
CURRENT APPLICATION NUMBER: US/10/140,002

CURRENT FILING DATE: 2002-05-06
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 543
LENGTH: 3721
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-002-543

Query Match 2.5%; Score 38.4; DB 4; Length 3721;
Best Local Similarity 47.0%; Pred. No. 5.9;
Matches 154; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 423 ACCCAGAGTCTGCTGGGGAGCTGTTTGGAGAGAGTGCTCTCAGAGAGGTCTG 482
DB 1887 ACCGTGGGCTGTGTTCTGTGGGCTCTCCGTGAGCGCGGAGCTGCGCCGGCGCAGGACAG 1926
QY 483 CACCCAGCC---CAGTGTGATGGGCGTCTTGTGCGCAGAGATGAGAGAGAG 539
DB 1927 GAGGCGATCTGGCGGAGCGGGGCGGAGCGGCTGAGCTCAGCGCTGCGGCGAG 1986
QY 540 TGTGCGGTGTGTCGGGACGGCGGAGAGCTATCTGTGAGCGGCTGCGGCGCTTC 599
DB 1987 CCGAGGGCGCAGAGGTCTCCGGGGCGGGGCGGAGCGGTGCGGTGCGCGCGGCGTTCC 2046
QY 600 CACTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCAAGTGGAGACTGGAGTCTCC 659
DB 2047 CCGAGGCTCTGTGCTGCGGCGCTCTGATGAGAAAGGCTGGGCGCAAGCGCTGTCAG 2106
QY 660 AGCTGCTGCGAGCAAGTCCAGAGAGTGAGCGCCCGGCGAGAGAGCGCCGCGCCAG 719
DB 2107 GCGGGGCGCCAGAGCTGAGCTGGGGGCTGTCGCGCAGCGCCGAGCAGCGCGCTGCGG 2166
QY 720 GAGCCACCCGCTGAGAGCCCGCTCCCGC 747
DB 2167 CAGAGCGCTGCGCCAGCTCCGAGCCCGC 2194

RESULT 66

US-09-191-171-7/C
Sequence 7, Application US/09191171
Patent No. 6149909
Patent No. 6149909 6143294
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,171
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995

APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8978Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-191-171-7

Query Match 2.5%; Score 38.4; DB 3; Length 4480;
Best Local Similarity 47.2%; Pred. No. 6.3;
Matches 150; Conservative 0; Mismatches 166; Indels 2; Gaps 1;

QY 592 GGGCCCTTCACCTGGGCTGCTGCTCCCTCCGCTCCGGAGATCCCAATGGAGACTGGA 651
DB 2618 GGAACGCCACCGTGGTGTCTGTCCAGACGATCCCGCTGCGACACTTCGGCCAGA 2559
QY 652 GGTGTCTCAGCTGCTGTCAGAGCAAGTCAGAGGTGAGCCCGGAGAGAGAGCCCC 711
DB 2558 GCTGCTCTCAT--CTGG 2501
QY 712 GGGCCCGAGAGCCACCGTGGAGAACCCCGCTCCCGGGGGCTTAAGTGGCGGAGAG 771
DB 2500 CGCCCCCGACCAAGGAGCGGGGCTCCCACTCGCTCCCTCCCTTGTGTGAAGAG 2441
QY 772 AGTAAGAGGTCCACTGGGGAACCCCTAGCGGCGATGACAGACTTGTGTACAAG 831
DB 2440 TCCCGAGCGCCCGCCAGCGAGAGCGAGCCCGGAGAGCGGGCTGAGAGAGTGGCG 2381
QY 832 ACCTGCGGCTCGCCTTCTGCAAGCCCGCTGCGAGATCTGAGCTCTGGCCCTGACC 891
DB 2380 TCCCGGAGCGGGGCTTTCAGAGAGCGCTCCCAAGCTGAGAGGGCCCTGGCGGCCA 2321
QY 892 CCTACTGTGTGGGTC 909
DB 2320 CCCAGCGGCGCCGCGCTC 2303

RESULT 67
US-09-385-707-7/c
Sequence 7, Application US/09385707
Patent No. 6238662
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOBWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDRONIDASE AND GENETIC
TITLE OF INVENTION: SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSES: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City

STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,707
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,104
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8978Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-385-707-7

Query Match 2.5%; Score 38.4; DB 3; Length 4480;
Best Local Similarity 47.2%; Pred. No. 6.3;
Matches 150; Conservative 0; Mismatches 166; Indels 2; Gaps 1;

QY 592 GGGCCCTTCACCTGGGCTGCTGCTCCCTCCGCTCCGGAGATCCCAATGGAGACTGGA 651
DB 2618 GGAACGCCACCGTGGTGTCTGTCCAGACGATCCCGCTGCGACACTTCGGCCAGA 2559
QY 652 GGTGTCTCAGCTGCTGTCAGAGCAAGTCAGAGGTGAGCCCGGAGAGAGAGCCCC 711
DB 2558 GCTGCTCTCAT--CTGG 2501
QY 712 GGGCCCGAGAGCCACCGTGGAGAACCCCGCTCCCGGGGGCTTAAGTGGCGGAGAG 771
DB 2500 CGCCCCCGACCAAGGAGCGGGGCTCCCACTCGCTCCCTCCCTTGTGTGAAGAG 2441
QY 772 AGTAAGAGGTCCACTGGGGAACCCCTAGCGGCGATGACAGACTTGTGTACAAG 831
DB 2440 TCCCGAGCGCCCGCCAGCGAGAGCGAGCCCGGAGAGCGGGCTGAGAGAGTGGCG 2381
QY 832 ACCTGCGGCTCGCCTTCTGCAAGCCCGCTGCGAGATCTGAGCTCTGGCCCTGACC 891
DB 2380 TCCCGGAGCGGGGCTTTCAGAGAGCGCTCCCAAGCTGAGAGGGCCCTGGCGGCCA 2321
QY 892 CCTACTGTGTGGGTC 909
DB 2320 CCCAGCGGCGCCGCGCTC 2303

RESULT 68
US-09-639-696C-6/c
Sequence 6, Application US/09639696C
Patent No. 6524835
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOBWOOD, John J.

TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDIURONIDASE AND GENETIC
SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON PEABODY LLP
STREET: 990 Stewart Avenue
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/639,696C
FILING DATE: 16-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/191,171
FILING DATE: 13-NOV-1998
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
APPLICATION NUMBER: AU PCT/AI92/00611
FILING DATE: 12-NOV-1992
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pokalsky, Ann R.
REGISTRATION NUMBER: 34,697
REFERENCE/DOCKET NUMBER: 2249/304
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-832-7572
TELEFAX: 516-832-7555
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 6238 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-639-696C-6
Query Match
Best Local Similarity 47.2%; Score 38.4; DB 4; Length 6238;
Matches 150; Conservative 0; Mismatches 166; Indels 2; Gaps 1;
QY 592 GGGCCCTCCAGCCGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGAGACTGGA 651
DB 4376 GGAACGCCACCGTGTGTTGCTGTCAGAGCGGTCCCGCTCCGACACTTCGGCCCA 4317
QY 652 GGTGCTCCAGCTGCTGCAGAGCAAGTCAGAGGTGAGCCCGGGAGAGAGAGCCCC 711
DB 4316 GCTGCTCTCAT--CTGCGGGGGGGGGGGGGGGGGCGCTGCGCGTGGGGTCTTGCCACG 4259
QY 712 GGGCCAGAGCCACCGTGGAGACCCGCTCCCGGGGGCTTAACTGCGCGGAGAG 771
DB 4258 CCGCCCGCCAGCCAGGGCGGGGCTCCACACTGCTCCCGCTCCCTTGGTGAAGAG 4199
QY 772 AGTAAAGAGTCCACTGGGGAACCCCTAGCCGGCATGACAGACTCTTGTACAAGC 831
DB 4198 TCCCCAGGGCCCGCACGGCGAGCTGACCCCGGAAGCGCGGTGAGAGAGTGGCC 4139
QY 832 ACTGCGGGCTCGGCTTCTGACAGCCCGCTGCGAGATCTGACTCTTGGCCCTGAC 891
DB 4138 TCCCCGACAGGGGCTTTCGACAGAGCCGCTCCCAAGCTGAGAGGCGCTTGGCCGCA 4079
QY 892 CCTACTGTGTGGGTC 909
||| |

DB 4078 CCCAGCGGCCCGGCTC 4061
RESULT 69
US-09-252-991A-6674
Sequence 6674, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6674
LENGTH: 1260
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6674
Query Match
Best Local Similarity 46.7%; Score 38.2; DB 4; Length 1260;
Matches 121; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
QY 974 GCTGGGCTGCTACTGCTGCGCGCTGCTCCACTGCGCGCTCCACTTCCAGCCGGCAC 1033
DB 114 GCTGGGCTGCTGACGCGCTTGGGGGGCCATGCGCCACCTCGGCTATTTCGGAGAGATT 173
QY 1034 CTCGGGCGCGGAGAGGGCTTGGCTGCAGATCTGCTAGAGAGACGTGACCCAGCCCC 1093
DB 174 CTTCCGCGCGCGCGCTGCTGAGGAAAGCAGCTACCGCGACTGCTGCTGCTGCGCA 233
QY 1094 TGTGAGGGGGTGTGCTGGCGCCCGAGCCCGCGCGCTGAGCCCTGGAGCCCAAGATGA 1153
DB 234 GTTCTTCCCGGCGCGCGCGAGAGGAGTGGCATGAGCTTGGCTGGCGCGCGCG 293
QY 1154 CACTGCAGTCAAGAGCCCGCTCTGTCACAGGATGACCTGAGTCCCTTGTAGGAGCA 1213
DB 294 CTATCCCGCGCGCTGCGCGCGCTGCTGCTGCTTCACTGCTGCGCTGCTGCTGCT 353
QY 1214 CACTTTCATGAGCATCTG 1232
DB 354 GCTCTTGGCTTCGCGCTG 372
RESULT 70
US-09-252-991A-6641
Sequence 6641, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6641
LENGTH: 2163
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6641
Query Match
2.5%; Score 38.2; DB 4; Length 2163;

Best Local Similarity 46.7%; Pred. No. 5.5;
Matches 121; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
QY 974 GCTGCGGTACTCACTGCGCCGCTTCCACTGCGCGTGCACCTTCCACCGGAC 1033
DB 847 GCTGCGGTACTCACTGCGCCGCTTCCACTGCGCGTGCACCTTCCACCGGAC 906
QY 1034 CTCGCGCGCGCGCGCGCGCTGCTGCGTGCAGACCTTCTGAGACGCGACCGCC 1093
DB 907 CGTCCCGCGCGCGCGCGCGCTGCTGCGTGCAGACCTTCTGAGACGCGACCGCC 966
QY 1094 TGTGAGGCGGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1153
DB 967 GTTCTCTCCCG 1026
QY 1154 CACTGCGAGTCAAGACCGCGCTGCGTGCAGAGTGAAGTGCCTTCTGAGCGAC 1213
DB 1027 CTATCCCG 1086
QY 1214 CACCTTGCATGCGATCTG 1232
DB 1087 GCTCTTGCCTCTGCGCTG 1105

RESULT 71
US-08-289-112-1
Sequence 1, Application US/08289112
Patent No. 5688640
GENERAL INFORMATION:
APPLICANT: Yamaoka, Masashi
TITLE OF INVENTION: Endothelin Converting Enzyme-1. A
TITLE OF INVENTION: Membrane-Bound Metalloprotease That Catalyzes The
NUMBER OF INVENTION: Proteolytic Activation of Big Endothelin-1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,112
FILING DATE: 10-Apr-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:414/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 118..2391
US-08-289-112-1

Query Match 2.5%; Score 38.2; DB 1; Length 2889;
Best Local Similarity 51.5%; Pred. No. 6.1;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 567 CTCATCTGTGTGACGCGCTGCGCGCTTCCACTGCGCGTGTGCTTCCGCTC 626
DB 157 CTCCTGGGAGAGGAGCGCGCTGCGCGTCAAGTCTTCCGCTGCGCTTCCGCTC 216
QY 627 CCGGAGATCCCAATGAGACCTGAGAGTCTTCAAGTCTTCCGCGCGTCAAGTCC 686
DB 217 CAGGTAACTTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 276
QY 687 GTGACCG 737
DB 277 GAGAACGCGCTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 327

RESULT 72
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 2.5%; Score 38.2; DB 3; Length 4403765;
Best Local Similarity 44.8%; Pred. No. 58;
Matches 187; Conservative 0; Mismatches 228; Indels 2; Gaps 1;
QY 681 CAGAGGTGACG 740
DB 3936918 CCG 3936859
QY 741 CTCCTCCCG 800
DB 3936858 GTGCG 3936799
QY 801 GCCGCGATGACAGACGCTTGTCTTCAAGACCTGCGCGCGCGCGCGCGCGCGCG 860
DB 3936798 TACACACG 3936739
QY 861 GTGCGAGTGTGACTCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 920
DB 3936738 CCG 3936679
QY 921 CAGAACCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 980
DB 3936678 GTGCG 3936619
QY 981 TGTACTCA--CTGCG 1038
DB 3936618 GTGCG 3936559
QY 1039 GCGCGCGAGCG 1095
DB 3936558 CCGCGAGCG 3936502

RESULT 73
US-08-458-568A-3/C
; Sequence 3, Application US/08458568A
; Patent No. 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
; TITLE OF INVENTION: Infections
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339r1s
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,568A
; FILING DATE: 02-JUNE-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,146
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn R.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: DFCI-0029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3439
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Herpes simplex virus
; STRAIN: Herpes Simplex Virus Type 1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..702
; US-08-458-568A-3

Query Match 2.5%; Score 38; DB 1; Length 702;
Best Local Similarity 50.5%; Pred. No. 4.2;
Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 989 CTGCGCGCTGCTTCACTGCGCTGCACTTCCAGCGCGACCTCCCGCGCGGAC 1048
DB 400 CAGAGGCGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 341
DB 1049 GGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108
DB 340 CGGGGGGGGGGGTGAACCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 281
QY 1109 GGGC 1168
DB 280 TGGCGCTGCG 221
QY 1169 GC 1170
DB 220 GC 219

RESULT 74
PCT-US91-06532-1
; Sequence 1, Application PC/TUS9106532
; GENERAL INFORMATION:
; APPLICANT: Rolzman, Bernard
; TITLE OF INVENTION: Recombinant Herpes Simplex Viruses
; TITLE OF INVENTION: Vaccines and Methods
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marchall, O'Toole, Gerstein, Murray &
; STREET: Two First National Plaza Suite 2100
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/06532
; FILING DATE: 19910910
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gruber, Lewis S.
; REGISTRATION NUMBER: 30,060
; REFERENCE/DOCKET NUMBER: 27373/8235
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/346-5750
; TELEFAX: 312/984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1335 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US91-06532-1

Query Match 2.5%; Score 38; DB 5; Length 1335;
Best Local Similarity 50.5%; Pred. No. 5.3;
Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 989 CTGCGCGCTGCTTCACTGCGCTGCACTTCCAGCGCGACCTCCCGCGCGGAC 1048
DB 407 CAGAGGCGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 466
QY 1049 GGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108
DB 467 CGGGGGGGGGGGTGAACCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 526
QY 1109 GGGC 1168
DB 527 TGGCGCTGCG 586
QY 1169 GC 1170
DB 587 GC 588

RESULT 75
US-08-196-218-31
; Sequence 31, Application US/08196218
; Patent No. 5614619
; GENERAL INFORMATION:
; APPLICANT: Piepersberg, Wolfgang
; APPLICANT: Stockmann, Michael
; APPLICANT: Taleghani, Kamiz Mansouri

APPLICANT: Dietler, Jurgen
APPLICANT: Grabley, Susanne
APPLICANT: Sichel, Petra
APPLICANT: Brnu, Barbara
TITLE OF INVENTION: Secondary-Metabolite Biosynthesis Genes
TITLE OF INVENTION: From Actinomycetes, Method of Isolating Them, and Their
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/196,218
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ogden, Scasia L.
REGISTRATION NUMBER: 36,228
REFERENCE/DOCKET NUMBER: 02481.1372-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 2634 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
FEATURE:
NAME/KEY: CDS
LOCATION: 416..1531
FEATURE:
NAME/KEY: CDS
LOCATION: 1561..2625
US-08-196-218-31
Query Match
Best Local Similarity 51.8%; Score 38; DB 1; Length 2634;
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 968 GGAGCGTGGCGGTGATCACTGCGCGCCCTGCTTCCATCGCGCTGCACCTTCCAC 1027
DB 2028 GGAGAAAGCGGTGACAGCGCGCGACGACATGCGCTGATCGGGTCTTACTTCCACCGC 2087
QY 1028 CGGCACTTCCCGCGCGCGGACGCGGCTGCGCTGACAGATCTGCTGAGAGACGTGACCCC 1087
DB 2088 CGCCATTCACGAGGGGGTGGCGGCCATCTCGCCGACAGCGCGGGGAACTGGAATCAG 2147
QY 1088 AGCCCTGTGAGAGGGGGTGGCGGCCCGCCAGCCCGCGCGCTGAGCC 1133
DB 2148 CGAGCGCGTTCAGTGGTGTGATCACTCGCGCGGAGCGTGGCGCC 2193
RESULT 76
US-08-681-953-31
Sequence 31, Application US/08681953
Patent No. 5710032
GENERAL INFORMATION:
APPLICANT: Plepersberg, Wolfgang

APPLICANT: Stockmann, Michael
APPLICANT: Taleghani, Kamaliz Mansouri
APPLICANT: Dietler, Jurgen
APPLICANT: Grabley, Susanne
APPLICANT: Sichel, Petra
APPLICANT: Brnu, Barbara
TITLE OF INVENTION: Secondary-Metabolite Biosynthesis Genes
TITLE OF INVENTION: From Actinomycetes, Method of Isolating Them, and Their
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,953
FILING DATE: 30-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/196,218
FILING DATE: 25-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ogden, Scasia L.
REGISTRATION NUMBER: 36,228
REFERENCE/DOCKET NUMBER: 02481.1372-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 2634 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
FEATURE:
NAME/KEY: CDS
LOCATION: 416..1531
FEATURE:
NAME/KEY: CDS
LOCATION: 1561..2625
US-08-681-953-31
Query Match
Best Local Similarity 51.8%; Score 38; DB 1; Length 2634;
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 968 GGAGCGTGGCGGTGATCACTGCGCGCCCTGCTTCCATCGCGCTGCACCTTCCAC 1027
DB 2028 GGAGAAAGCGGTGACAGCGCGCGACGACATGCGCTGATCGGGTCTTACTTCCACCGC 2087
QY 1028 CGGCACTTCCCGCGCGCGGACGCGGCTGCGCTGACAGATCTGCTGAGAGACGTGACCCC 1087
DB 2088 CGCCATTCACGAGGGGGTGGCGGCCATCTCGCCGACAGCGCGGGGAACTGGAATCAG 2147
QY 1088 AGCCCTGTGAGAGGGGGTGGCGGCCCGCCAGCCCGCGCGCTGAGCC 1133
DB 2148 CGAGCGCGTTCAGTGGTGTGATCACTCGCGCGGAGCGTGGCGCC 2193
RESULT 77


```
QY 500 CATGGGCGTCTCTTGGAGAGAGTGGCTCTCAGAGAGGAGTGGCTTGTGTGCGGACCG 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9951 GGAGGACGAGATCCGGGTGTGCTGTGAGACGAGGAGCGGTACCGGACGTTCTGGACGT 10010
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 560 CGGGAGCTCATCTGCTGTGAGCGGCTGCCCTCGGGCTTCCACTTGACC 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 10011 CGCCCGCTCAACGAGCCCGGAGCGCCGCTCTGTCTCCGGGACGCGGAC 10059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 80
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match
Best Local Similarity 51.5%; Pred. No. 18;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 440 GGGAGCTCTTTTGGAGAGAGTGGCTCTCAGAGAGGAGTGGCTTGTGTGCGGACCG 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11632 GGGCGCGCGCTCATGACGAGAGTGGCTGCCCGCGGTGGCGGAGTCTGCGCTCCAGGCGCG 11691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 500 CATGGCGCTCTTGTGCTGTGCCAAGATGAGAGCGAGTGGCTGTGTGCGGACCG 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11692 GGAGGACGAGATCCGGGTGTGCTGTGAGACGAGGAGCGGTACGGCGGACGTTCTGGACGT 11751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 560 CGGGAGCTCATCTGCTGTGAGCGGCTGCCCTCGGGCTTCCACTTGACC 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11752 CGCCCGCTCAACGAGCCCGGAGCGCCGCTCTGTCTCCGGGACGCGGAC 11800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 81
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match
Best Local Similarity 51.5%; Pred. No. 18;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 440 GGGAGCTCTTTTGGAGAGAGTGGCTCTCAGAGAGGAGTGGCTTGTGTGCGGACCG 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9774 GGGCGCGCGCTCATGACGAGAGTGGCTGCCCGCGGTGGCGGAGTCTGCGCTCCAGGCGCG 9833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 500 CATGGCGCTCTTGTGCTGTGCCAAGATGAGAGCGAGTGGCTGTGTGCGGACCG 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9834 GGAGGACGAGATCCGGGTGTGCTGTGAGACGAGGAGCGGTACGGCGGACGTTCTGGACGT 9893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 560 CGGGAGCTCATCTGCTGTGAGCGGCTGCCCTCGGGCTTCCACTTGACC 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9894 CGCCCGCTCAACGAGCCCGGAGCGCCGCTCTGTCTCCGGGACGCGGAC 9942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match
Best Local Similarity 51.5%; Pred. No. 18;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 440 GGGAGCTCTTTTGGAGAGAGTGGCTCTCAGAGAGGAGTGGCTTGTGTGCGGACCG 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9774 GGGCGCGCGCTCATGACGAGAGTGGCTGCCCGCGGTGGCGGAGTCTGCGCTCCAGGCGCG 9833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 500 CATGGCGCTCTTGTGCTGTGCCAAGATGAGAGCGAGTGGCTGTGTGCGGACCG 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9834 GGAGGACGAGATCCGGGTGTGCTGTGAGACGAGGAGCGGTACGGCGGACGTTCTGGACGT 9893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 560 CGGGAGCTCATCTGCTGTGAGCGGCTGCCCTCGGGCTTCCACTTGACC 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9894 CGCCCGCTCAACGAGCCCGGAGCGCCGCTCTGTCTCCGGGACGCGGAC 9942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 82
US-09-141-908-1
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-1

Query Match
Best Local Similarity 51.5%; Pred. No. 18;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 440 GGGAGCTCTTTTGGAGAGAGTGGCTCTCAGAGAGGAGTGGCTTGTGTGCGGACCG 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9774 GGGCGCGCGCTCATGACGAGAGTGGCTGCCCGCGGTGGCGGAGTCTGCGCTCCAGGCGCG 9833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 500 CATGGCGCTCTTGTGCTGTGCCAAGATGAGAGCGAGTGGCTGTGTGCGGACCG 559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9834 GGAGGACGAGATCCGGGTGTGCTGTGAGACGAGGAGCGGTACGGCGGACGTTCTGGACGT 9893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 560 CGGGAGCTCATCTGCTGTGAGCGGCTGCCCTCGGGCTTCCACTTGACC 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9894 CGCCCGCTCAACGAGCCCGGAGCGCCGCTCTGTCTCCGGGACGCGGAC 9942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 83

US-09-657-440-19
; Sequence 19, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-657-440-19

Query Match 2.4%; Score 37.8; DB 4; Length 38506;
Best Local Similarity 51.5%; Pred. No. 18;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 440 GGGGAGCTTTTGGAGAGAGGTGCTCTCAGAGAGGTGTCGACCCAGCCCACTCTG 499
DB 9774 GGGCGCCCGGCTATCAGAGAGTGCCTCCCGGCTGCGGAGCTCCAGGCC 9833
QY 500 CATGGCGTCTCTTGGCTTGTCCAGAGATGAGACGAGTGTGCGTGTGCGGAGCG 559
DB 9834 GAGGAGCGAGATCCGCTGTGCTGAGACGAGAGCCGTAACGCGGACGTCTGACGT 9893
QY 560 CGGGAGCTCATCTGCTGTGAGCGGCTCCCTCGGGCTTCCACTGTCC 608
DB 9894 CGCGCGCTCAGCGGCCGAGCGCGCTCTGTCCGGGACGCGGAC 9942

RESULT 84

US-09-305-640-1/C
; Sequence 1, Application US/09305640B
; Patent No. 6255468
; GENERAL INFORMATION:
; APPLICANT: SmithKline Beecham plc
; TITLE OF INVENTION: No. 6255468el Compounds
; FILE REFERENCE: GP30124
; CURRENT APPLICATION NUMBER: US/09/305,640B
; CURRENT FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2560
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-640-1

Query Match 2.4%; Score 37.6; DB 3; Length 2560;
Best Local Similarity 46.8%; Pred. No. 8.3; Indels 0; Gaps 0;
Matches 118; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 397 CAGGCGAGAGACTGGGAGATTGACGATCCAGAGATCTGCTGGGAGACTGTTTGGGA 456
DB 514 CGGAGACAGCG 455
QY 457 AGGAGTGTCTCTCAGAGAGGTGCTGACCCCGAGCTTCTGATGAGGAGTCTTTGCC 516
DB 454 CCAGATGGCGGAGAGCG 395

RESULT 85

US-10-140-002-119/C
; Sequence 119, Application US/10140002
; Patent No. 6725730
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroli, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zeman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C59
; CURRENT APPLICATION NUMBER: US/10/140,002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 119
; LENGTH: 2868
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-002-119

Query Match 2.4%; Score 37.6; DB 4; Length 2868;
Best Local Similarity 46.8%; Pred. No. 8.6;
Matches 118; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 397 CAGGCGAGAGACTGGGAGATTGACGATCCAGAGATGCTGCTGGGAGACTGTTTGGGA 456
DB 494 CGGAGACAGCG 435
QY 457 AGGAGTGTCTCTCAGAGAGGTGCTGACCCCGAGCGATCTGATGAGGAGCTCTTGGCC 516
DB 434 CCAGATGGCGGAGAGCG 375
QY 517 TGTGCCAGAGAGATGAGACGAGTGTGCTGTGTCGAGACCGCGGAGAGCTCATCTGCT 576
DB 374 GGTTCAGAGGCGGAGCG 315
QY 577 GTGAGCGCTGCCCTTCCAGCTTTCACCTGAGCTGCTGCTTCCCTCCCTCCGCGAGATCC 636
DB 314 AGCCCGGCGGAGAGAGCG 255
QY 637 CCAGTGGAGACT 648
DB 254 CCTTTGAACT 243

RESULT 86
US-09-774-528-383/c
Sequence 383, Application US/09774528
Patent No. 6743619
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhou, Ping
APPLICANT: Goodrich, Ryle
APPLICANT: Liu, Chenshua
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyun
APPLICANT: Zhang, Jie
APPLICANT: Zhao, Qing A.
APPLICANT: Yang, Yonghong
APPLICANT: Xue, Aidong J.
APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Wang, Dunrui
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6743619el Nucleic Acids and
FILE REFERENCE: 802
CURRENT APPLICATION NUMBER: US/09/774,528
CURRENT FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 441
SOFTWARE: pc_fl_genes Version 2.0
SEQ ID NO 383
LENGTH: 1706
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (92)..(1309)
US-09-774-528-383

Query Match 2.4%; Score 37.4; DB 4; Length 1706;
Best Local Similarity 50.9%; Pred. No. 8.1;
Matches 89; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 627 CGGAGATCCCGAGGAGCTGAGGCTCCAGCTGCTCGAGCAACAGCCAGAG 686
DB 1199 CGCGAGGCCCGAGAGGGGATGGGGCTTCGAGGGGGCTGGGCTCATGAGAAAGTG 1140
QY 687 GTGCAAGCCCGGAGAGAGAGCCCGGCGCCAGAGCCACCCGTGAGACCCCGCTCCC 746
DB 1139 CTGAAGCTTCCTCGAGAGGGTGCCAGCGAGAGAGGCCACCTGACACTATCTGGGGCC 1080
QY 747 CGGGGGCTTAGCTCGCGGGAGAGAGGTAAAGGTCCACTGGGGAGAACCCCTAG 801
DB 1079 CAGAGGCTGAGACGAGCGCTCTGAGTCAAGAGAGGTGGCTGAGAGCCCTGG 1025

RESULT 87
US-09-252-991A-9715/c
Sequence 9715, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfeld et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 9715
LENGTH: 648
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-9715

Query Match 2.4%; Score 37.2; DB 4; Length 648;
Best Local Similarity 46.5%; Pred. No. 6.5;
Matches 120; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 409 TGGGAGTTCAAGTACCCAGAGATGCTGTGGGGAGAGCTGTTTGGGAAGAGAGTGCGTC 468
DB 463 TGGGGCTCAGCGGCTTGGCCGGATGCCCTGGGGTGAAGGCCGGGGTGAAGAGGTCCT 404
QY 469 TCAGAGAGGTGCTGACCCAGCCAGCCAGCTGATGGAGTCTCTTGGCTGTGCCAAGA 528
DB 403 GTCCGAGGCCATGATGACTTCAGCGCGGCTTCAAGTCTCAGCGGGGCTGGCTGCA 344
QY 529 ATGAGACGAGTGTGCGGTGTGTGCGGACGCGGGAGACTATCTGTGTGACGCGTGC 588
DB 343 TGTAGCCGAGGCGCGGCTGCGCAGACCGCTTGGCCAGCACGCGCGCTGAACACCG 284
QY 589 CTCGGGCTTCCACCTGGCGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGAGCT 648
DB 283 AGTAGACCATCACTTCAATTGGCGGGCAGTCCGGGTCAAGCATTCACAGGTGATAC 224
QY 649 GAGGTGCTCCAGCTGCC 666
DB 223 CCTCGCGCTGCGCTGCC 206

RESULT 88
US-08-318-837-1
Sequence 1, Application US/08318837
Patent No. 5981277
GENERAL INFORMATION:
APPLICANT: FRANSSEN, LUCIA; DEVOS, KATHLEEN; VAN DE VOORDE,
APPLICANT: ANDRE; VAN HEUVERSWIJN, HUGO
TITLE OF INVENTION: NEW POLYPEPTIDES AND PEPTIDES, NUCLEIC ACID
TITLE OF INVENTION: COOKING FOR THEM, AND THEIR USE IN THE FIELD OF TUMOR THERAPY
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN AND MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,837
FILING DATE: 13-OCT-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP 93/01022
FILING DATE: 28-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 92,401,231.3
FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 410,007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2150 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Human
TISSUE TYPE: Spleen tissue from healthy human
FEATURE:
NAME/KEY: CDS
LOCATION: 1810..1982
FEATURE:
NAME/KEY: misc_feature
LOCATION: 543
OTHER INFORMATION: C may be present or absent
FEATURE:
NAME/KEY: Intron
LOCATION: 1980..2150
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..1809
US-08-318-837-1

Query Match 2.4%; Score 37.2; DB 2; Length 2150;
Best Local Similarity 51.2%; Pred. No. 9.8;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 584 CTGCCCTCGGCGCTTCACCTGGCGCTGCTCCCTCCGCGGAGATCCCGAGTGG 643
Db 1709 CCGGACTTCAAGCCCGCCCGCCCGCCGCTCCCGGCTCGGAGTGTCTCCGGGG 1768
Qy 644 GACCTGAGAGTGTCTCCAGCTCCCTGAGGCAACATGTCAGAGAGTCCAGCCCGGGCAGA 703
Db 1769 GTCCGCGAGCGCGGGCGCGGCGCGAGCGCGGCGGAGCATGCGGGCGCGCGCGGG 1828
Qy 704 GGAGCCCCCGCCCGAGAGCCCGTGGAGAGCCCGGCTCCCGCGGGGC 753
Db 1829 CGGCTGGGCGCGCGGCGGCGAGCGTGGCGCGGAGCCCGCCCGCGGGC 1878

RESULT 89
US-09-428-711A-13/C
Sequence 13, Application US/09428711A
Patent No. 6358720
GENERAL INFORMATION:
APPLICANT: Muramatsu, Masaaki
APPLICANT: Shirasawa, Takuji
APPLICANT: Tokumitsu, Hiroshi
APPLICANT: NO. 6358720uchl, Teruhisa
TITLE OF INVENTION: SERINE/THREONINE PROTEIN KINASE
FILE REFERENCE: 06501-045001
CURRENT APPLICATION NUMBER: US/09/428,711A
CURRENT FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: PCT/JP98/01246
PRIOR FILING DATE: 1998-03-23
PRIOR APPLICATION NUMBER: JP 9/124798
PRIOR FILING DATE: 1997-04-28
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 3995
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (388) ... (3540)
NAME/KEY: misc_feature
LOCATION: (1) ... (3995)
OTHER INFORMATION: n = A,T,C or G
US-09-428-711A-13

Query Match 2.4%; Score 37.2; DB 3; Length 3995;
Best Local Similarity 46.2%; Pred. No. 12;
Matches 117; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 890 CCCCTACTGTGTGTGGTCTGAGGATGACGAGAACTGGCTCTGGGCGCGTGGCGG 949
Db 393 CTCATATAGCGCGCGGGCGCGGAGCGGCGCTCGGCGCCCGGGCGGCTACCG 334
Qy 950 GGTGTGCGGAGATGTGACGAGAGTGTGTGTACTCACTGCGCGCTGCTTCACTG 1009
Db 333 GGTGCGTGGCTCAGCGGAGCGGCGCGGCGCGCTGACGAGGACGAGCGCGGCTC 274
Qy 1010 GGGCTGCCACTTCCAGCGGAGACTCCCGCGCGGAGCGGCGCTGCGAGATCTCG 1069
Db 273 ATGCGAGGCGCGCTGCTCCCATCGGAGCGCTAGCGCGGATCCAGTCCACAGCGCG 214
Qy 1070 CTGAGAGACGTGACCCGAGCCCTGTGTGAGGAGGAGTGTGCGCCCGCGCGCGCT 1129
Db 213 GGGAGCAGAGGCGCGCTCCTCGCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 154
Qy 1130 GGGCGCTGGGCT 1142
Db 153 GGACCGGTGACT 141

RESULT 90
US-09-418-710-28
Sequence 28, Application US/09418710
Patent No. 6596482
GENERAL INFORMATION:
APPLICANT: Jones, Michael H.
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
FILE REFERENCE: 06501-042001
CURRENT APPLICATION NUMBER: US/09/418,710
CURRENT FILING DATE: 1999-10-15
PRIOR APPLICATION NUMBER: PCT/JP98/01783
PRIOR FILING DATE: 1998-04-17
PRIOR APPLICATION NUMBER: JP 9/310027
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: JP 9/116570
PRIOR FILING DATE: 1997-04-18
NUMBER OF SEQ ID NOS: 73
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 5561
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (346) ... (4926)
US-09-418-710-28

Query Match 2.4%; Score 37.2; DB 4; Length 5561;
Best Local Similarity 53.4%; Pred. No. 13;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 535 ACGAGTGCCTGTGTGTGCGGACGCGGAGAGCTATCTGCTGTACGCGTCCCTCGGG 594
Db 3908 AAGTTGTTCAGAAAGGTGAGTGAACAATTATCTTGTGTATGATGATTAAG 3967
Qy 595 CTTCCACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 654
Db 3968 CTTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4027
Qy 655 GCTCCAGCTGCTGCGAGGCAACAGTC 680
Db 4028 GCCGAGCTTGCAGCGCGCTACTGCC 4053

RESULT 91
US-09-839-479-28
Sequence 28, Application US/09839479
Patent No. 6727222
GENERAL INFORMATION:
APPLICANT: Jones, Michael H.
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
FILE REFERENCE: 06501-042002


```
;; CURRENT APPLICATION NUMBER: US/09/839,479
;; CURRENT FILING DATE: 2001-04-20
;; PRIOR APPLICATION NUMBER: US 09/418,710
;; PRIOR FILING DATE: 1999-10-15
;; PRIOR APPLICATION NUMBER: PCT/JP98/01783
;; PRIOR FILING DATE: 1998-04-17
;; PRIOR APPLICATION NUMBER: JP 9/310027
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: JP 9/116570
;; PRIOR FILING DATE: 1997-04-18
;; NUMBER OF SEQ ID NOS: 72
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 28
;; LENGTH: 5561
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (346)...(4926)
US-09-839-479-28
```

```
Query Match
Best Local Similarity 53.4%; Score 37.2; DB 4; Length 5561;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
```

```
QY 535 ACGAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGCTGACGGCTGCGCGG 594
DB 3908 AAGTTGTCCAAAGAAAGGTGAGATGACAAATTGATCTGTGTGATGATTAATAAAG 3967
QY 595 CCTTCACCTGCGCTGCTGCTCCCTCGCTCGGAGATCCCAATGGAGACCTGAGGT 654
DB 3968 CCTTCACCTGTTTGTCTGAGGCGCGCCCTCTATGATGACCAATGGAGTGCAGT 4027
QY 655 GCTCCAGCTGCTGTCAGGCAACAGTC 680
DB 4028 GCCCAGCTTGCCAGCCCGGCTACTGCC 4053
```

```
RESULT 92
US-09-418-710-30
; Sequence 30, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 5573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (346)...(4938)
US-09-418-710-30
```

```
Query Match
Best Local Similarity 53.4%; Score 37.2; DB 4; Length 5573;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
```

```
QY 535 ACGAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGCTGACGGCTGCGCGG 594
DB 3920 AAGTTGTCCAAAGAAAGGTGAGATGACAAATTGATCTGTGTGATGATTAATAAAG 3979
```

```
QY 595 CCTTCACCTGCGCTGCTGCTCCCTCGCTCGGAGATCCCAATGGAGACCTGAGGT 654
DB 3980 CCTTCACCTGTTTGTCTGAGGCGCGCCCTCTATGATGACCAATGGATGAGTGCAGT 4039
QY 655 GCTCCAGCTGCTGTCAGGCAACAGTC 680
DB 4040 GCCCAGCTTGCCAGCCCGGCTACTGCC 4065
```

```
RESULT 93
US-09-839-479-30
; Sequence 30, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 5573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (346)...(4938)
US-09-839-479-30
```

```
Query Match
Best Local Similarity 53.4%; Score 37.2; DB 4; Length 5573;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
```

```
QY 535 ACGAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGCTGACGGCTGCGCGG 594
DB 3920 AAGTTGTCCAAAGAAAGGTGAGATGACAAATTGATCTGTGTGATGATTAATAAAG 3979
QY 595 CCTTCACCTGCGCTGCTGCTCCCTCGCTCGGAGATCCCAATGGAGACCTGAGGT 654
DB 3980 CCTTCACCTGTTTGTCTGAGGCGCGCCCTCTATGATGACCAATGGATGAGTGCAGT 4039
QY 655 GCTCCAGCTGCTGTCAGGCAACAGTC 680
DB 4040 GCCCAGCTTGCCAGCCCGGCTACTGCC 4065
```

```
RESULT 94
US-08-146-930-1/c
; Sequence 1, Application US/08146930
; Patent No. 5958764
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothenberg, Joseph A.
; APPLICANT: Greenhalgh, David A.
; TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
```

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/146,930
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application two
APPLICATION NUMBER: 07/876,286
FILING DATE: April 30, 1992
APPLICATION NUMBER: No. 5958764 yet assigned (204/132)
FILING DATE: October 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/152
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6530 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-146-930-1

Query Match 2.4%; Score 37.2; DB 2; Length 6530;
Best Local Similarity 47.8%; Pred. No. 14;
Matches 149; Conservative 0; Mismatches 153; Indels 10; Gaps 1;
QY 581 CGGCTGCTCCGCGGCTTCCACCTGCGCTGCTCCGCGCTCCGCGGAGATCCCGAG 640
DB 3487 CCGCGCCACCGAGAGAACCTTCCGACGCTGTCGCCCGCACCGAATAGCCGCCG 3428
QY 641 TGGGACCTGAGAGTGTCTCCAGCTGCTGAGGCAACATGTCAGAGAGTCCAGCCCGGCG 700
DB 3427 CCGCGCGCGACGTGATCCACCGCACCTG-----AGTACCGCGCGCACAGCT 3378
QY 701 AAGAGAGCCCCCGGCCCCAGAGAGCCAGTCCGCGCTCCCCCGGAGCTTAGTTC 760
DB 3377 AAGAGCGCGCTGCTTCCACCGCGCGGAAATAGCGCTCCGACAGTGAAGCCCGCC 3318
QY 761 GCGGGAGAGAGAGTGAAGAGTCCACCTGGGGAACCCCTAGCGGAGATGAGACAGACTCT 820
DB 3317 GCCACCAAGTACTTCCCGCTCCGAAACCGCGCGGACCAAGGAGTGCCTGCTTC 3258
QY 821 TGTCTAACAAGACCTGCGGCTCCGCTTCTGCAAGCCCGCTGCAAGTCTGACTCTTC 880
DB 3257 AAGAGAGCTCTCGCAGTAGAGAGCGCGCTCTGATCCGCTCCGACAGTGAAGCGCGC 3198
QY 881 GGCCTTGACACC 892
DB 3197 GCCGCGCGCCCC 3186

RESULT 95
US-08-458-240-1/c
; Sequence 1, Application US/08458240
; Patent No. 6143727
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothenberg, Joseph A.
; APPLICANT: Greenhalgh, David A.
; TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
; NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,240
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/146,930
FILING DATE:
APPLICATION NUMBER: 07/876,286
FILING DATE: April 30, 1992
APPLICATION NUMBER: No. 6143727 yet assigned (204/132)
FILING DATE: October 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/152
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6530 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-458-240-1

Query Match 2.4%; Score 37.2; DB 3; Length 6530;
Best Local Similarity 47.8%; Pred. No. 14;
Matches 149; Conservative 0; Mismatches 153; Indels 10; Gaps 1;
QY 581 CGGCTGCTCCGCGGCTTCCACCTGCGCTGCTCCGCGCTCCGCGGAGATCCCGAG 640
DB 3487 CCGCGCCACCGAGAGAACCTTCCGACGCTGTCGCCCGCACCGAATAGCCGCCG 3428
QY 641 TGGGACCTGAGAGTGTCTCCAGCTGCTGAGGCAACATGTCAGAGAGTCCAGCCCGGCG 700
DB 3427 CCGCGCGCGACGTGATCCACCGCACCTG-----AGTACCGCGCGCACAGCT 3378
QY 701 AAGAGAGCCCCCGGCCCCAGAGAGCCAGTGAAGACCCGCTCCCCCGGAGCTTAGTTC 760
DB 3377 AAGAGCGCGCTGCTTCCACCGCGCGGAAATAGCGCTCCGACAGTGAAGCCCGCC 3318
QY 761 GCGGGAGAGAGAGTGAAGAGTCCACCTGGGGAACCCCTAGCGGAGATGAGACAGACTCT 820
DB 3317 GCCACCAAGTACTTCCCGCTCCGAAACCGCGCGGACCAAGGAGTGCCTGCTTC 3258
QY 821 TGTCTAACAAGACCTGCGGCTCCGCTTCTGCAAGCCCGCTGCAAGTCTGACTCTTC 880
DB 3257 AAGAGAGCTCTCGCAGTAGAGAGCGCGCTCTGATCCGCTCCGACAGTGAAGCGCGC 3198
QY 881 GGCCTTGACACC 892
DB 3197 GCCGCGCGCCCC 3186

RESULT 96
PCT-US93-03993-1/c
; Sequence 1, Application PC/TUS9303993
; GENERAL INFORMATION:

QY 1095 GTGAGAGGGGTGCTGGCCCCCAGCCCCCGCTGAGCCCCCTGAGCTGCC 1145
DB 114 CGGCGTTCCGCGCAGCTGCTGGCGCGCTCTGTCGGGTCCGGGCGCAGCC 64

RESULT 98

US-09-022-765-34/C
Sequence 34, Application US/09022765
Patent No. 6375955
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yahir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,765
FILING DATE: 12-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLSCULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Leishmania major
US-09-022-765-34

Query Match 2.4%; Score 37; DB 3; Length 516;
Best Local Similarity 44.8%; Pred. No. 6.8;
Matches 184; Conservative 0; Mismatches 225; Indels 2; Gaps 1;

QY 737 CCCGCTCCCCCGGGCTTAAAGTCCGCGGAGAGAGGTAAGAGTCCAGCTGGGGAACC 796
DB 474 CGGCTGCTGCGTGGCGGTGTCTAGGCGCTGCTGAGCTCTCGGGGTTCCGGGCGCAGCT 415
QY 797 CCTAGCCGCGCATGACACGACTTGTCTAACAGCAGCTGCGGCTCCGCTTCTGACG 856
DB 414 GCTGGGCGGCTGTCGCGGTCCGGGCGCAGCGGTCCAGCCGCTGCTCAGCTCGGCGC 355
QY 857 CCCGCTGCGAGTCTGACTCTCGGCGCTGCAACCCCTACTGTGTGTGAGTCTGAGG 916
DB 354 GCTGCTGCGTGGCGGTGTCTAGGCGCTGCTGAGCTCTCGGGTTCCGGGCGCAGCGGTG 295
QY 917 TCAGCAGAACT--GGCTCTGCTGCGGCGGTGCGGGGTGCGGAGATGTACGAGCGTG 974
DB 294 CCACCTGTGCTCCAGCTGCGGCGGCTGCTGCGGTGTCTAGGCGCTGTGAGCT 235
QY 975 CTGCGGTACTCACTGCGCGGCTGCTTCACTGCGGCTGCACTTCCAGCGGCGACC 1034

DB 234 CTTCCGCGTTCCGCGCAGCTGCTGAGCGGCGCTGCTCGGCGGCGCAGCTGCA 175
QY 1035 TCCCGCGCGGAGCGGCGCTGCGTGCAGATCTGCTCAGAGAGCTGACCCAGCCCT 1094
DB 174 CCGGTGCTCCAGCTCGGCGGCGCTGCTGAGCGGTGTCTAGGCGCTGCTGCACTGCT 115

QY 1095 GTGAGAGGGGTGCTGGCCCCCAGCCCCCGCTGAGCCCCCTGAGCTGCC 1145
DB 114 CGGCGTTCCGCGCAGCTGCTGGCGCGCTCTGTCGGGTCCGGGCGCAGCC 64

RESULT 99

US-09-551-974A-34/C
Sequence 34, Application US/09551974A
Patent No. 6500437
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yahir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
FILE REFERENCE: 210121.420C5
CURRENT APPLICATION NUMBER: US/09/551,974A
NUMBER OF SEQ ID NOS: 101
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 34
LENGTH: 516
TYPE: DNA
ORGANISM: Leishmania major
US-09-551-974A-34

Query Match 2.4%; Score 37; DB 4; Length 516;
Best Local Similarity 44.8%; Pred. No. 6.8;
Matches 184; Conservative 0; Mismatches 225; Indels 2; Gaps 1;

QY 737 CCCGCTCCCCCGGGCTTAAAGTCCGCGGAGAGAGGTAAGAGTCCAGCTGGGGAACC 796
DB 474 CGGCTGCTGCGTGGCGGTGTCTAGGCGCTGCTGAGCTCTCGGGGTTCCGGGCGCAGCT 415
QY 797 CCTAGCCGCGCATGACACGACTTGTCTAACAGCAGCTGCGGCTCCGCTTCTGACG 856
DB 414 GCTGGGCGGCTGTCGCGGTCCGGGCGCAGCGGTCCAGCCGCTGCTCAGCTCGGCGC 355
QY 857 CCCGCTGCGAGTCTGACTCTCGGCGCTGCAACCCCTACTGTGTGTGAGTCTGAGG 916
DB 354 GCTGCTGCGTGGCGGTGTCTAGGCGCTGCTGAGCTCTCGGGTTCCGGGCGCAGCGGTG 295
QY 917 TCAGCAGAACT--GGCTCTGCTGCGGCGGTGCGGGGTGCGGAGATGTACGAGCTG 974
DB 294 CCACCTGTGCTCCAGCTGCGGCGGCTGCTGCGGTGTCTAGGCGCTGTGAGCT 235
QY 975 CTGCGGTACTCACTGCGCGGCTGCTTCACTGCGGCGCTGCACTTCCAGCGGCGACC 1034
DB 234 CTTCCGCGTTCCGCGCAGCTGCTGAGCGGCGCTGCTGAGCGGTGTCTAGGCGCTGTGAGCT 175
QY 1035 TCCCGCGCGGAGCGGCGCTGCGTGCAGATCTGCTCAGAGAGCTGACCCAGCCCT 1094
DB 174 CCGGTGCTCCAGCTCGGCGGCGCTGCTGAGCGGTGTCTAGGCGCTGTGAGCT 115
QY 1095 GTGAGAGGGGTGCTGGCCCCCAGCCCCCGCTGAGCCCCCTGAGCTGCC 1145
DB 114 CGGCGTTCCGCGCAGCTGCTGGCGCGCTCTGTCGGGTCCGGGCGCAGCC 64

RESULT 100

US-09-565-501A-34/C
Sequence 34, Application US/09565501A
Patent No. 6607731
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Bhatia, Ajay
APPLICANT: Coler, Rhea
APPLICANT: Peter Prober
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
THERAPY AND DIAGNOSIS OF LEISHMANIASIS
FILE REFERENCE: 210121.420C6
CURRENT APPLICATION NUMBER: US/09/565,501A
NUMBER OF SEQ ID NOS: 112
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 34
LENGTH: 516
TYPE: DNA
ORGANISM: Leishmania major
US-09-565-501A-34

Query Match 2.4%; Score 37; DB 4; Length 516;
Best Local Similarity 44.8%; Pred. No. 6.8;
Matches 184; Conservative 0; Mismatches 225; Indels 2; Gaps 1;

QY 737 CCCGCTCCCCCGGGCTTAAAGTCCGGGAGAGAGTAAAGTTCACCTGGGGAACC 796
DB 474 CCGGCTGCTGCGTGGGCGGTCTAAGGCGCTGCTGCACTCTCGCGCTTCGCGCCAGCT 415
QY 797 CCTAGCCGAGATGAGACAGACTTGTCTAAGACACTGCGGCTCCGCTTCTGACG 856
DB 414 GCTGGGCGCTGCTGCGGCTCGCGGCAAGCCGTGCAACCGTCTCAAGCTCGGCGC 355
QY 857 CCGGCTGCCAGGCTGGAATCTCTGAGCCCTGCAACCTTACTGTGTGTGAGGAG 916
DB 354 GCTGCTGCGGTGGGAGTGTCTAAGCGCTGCTGCACTCTCGCGGCTTCGCGGCAAGCCGTG 295
QY 917 TCAAGCAAGACT--GGCTCTGTGCGCGGCTTGGCGGAGTGTCCGAGATGTACGACGTG 974
DB 294 CCACCTGTGCTTCAAGCTGCGGCGCGCTGCTGCGGTGCTGAGCGGTGCTGCAAGCT 235
QY 975 CTGCGGTGTAATCACTGAGCGCGCTTCACTGAGCGTGCACCTTCCAGCCGAGCACC 1034
DB 234 CTTGCGGCTTCCGCGGCAAGCTGCTGAGCGGCTCTGTCGCGCTTCGCGGCAAGCCGTGCA 175
QY 1035 TCCCGGCCGAGAGAGCGGCTGCGCTGCAAGTCTGCTCAAGAGACGTGAACCCAGCCCT 1094
DB 174 CCGGTGCTTCAAGCTGCGGCGCGCTGCTGAGCGGTGTAGAGCGCTGTGCAAGCTCT 115
QY 1095 GTGAGAGGAGTGTGAGCGGCGGCAAGCCCGCGCTGAGCCCTGAGGCTTGGC 1145
DB 114 CGGCGTTTCGCGGCAAGCTGCTGAGCGGCGCTGCTGCGGCTTCGCGGCAAGCC 64

Search completed: February 9, 2005, 14:36:14
Job time : 195 sec

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 09:10:56 ; Search time 4527 Seconds
(without alignments)
12436.356 Million cell updates/sec

Title: US-09-508-658a-3

Perfect score: 1545
Sequence: 1 agagaagaagtgaggcttcctc.....aaaataataaataagtcgc 1545

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 200 summaries

Database :

EST: *
1: gb_esc1: *
2: gb_esc2: *
3: gb_hcc: *
4: gb_esc3: *
5: gb_esc4: *
6: gb_esc5: *
7: gb_esc6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	612.4	39.6	642	5	BX112453 BX112453
2	542.6	35.1	1506	9	AY419550 Homo sapi
3	400	25.9	1474	9	AY419551 Pan trogl
4	374.6	24.2	433	1	AA742555 nx30c11.8
5	283.4	18.3	1512	9	AY419552 Mus muscu
6	162.6	10.5	451	2	BE627856 u49c07.y
7	156.2	10.1	632	6	BY728790 BY728790
8	149.4	9.7	1399	3	BC068026 Mus muscu
9	111.4	7.2	511	2	BE630816 u49c07.x
10	91.4	5.9	542	1	AA866822 vx91b07.r
11	90.2	5.8	420	1	CG513175 OSTR6421
12	86	5.6	384	1	AT552580 vx26b07.x
13	73.4	4.8	947	9	AG069825 Pan trogl
14	73	4.7	959	9	CNS033FO Droso
15	69.6	4.5	925	9	AL053013 Droso
16	69.4	4.5	925	4	BM740688 UI-E-EJO-
17	69.4	4.5	250	4	BM740688 UI-E-EJO-
18	67	4.3	940	7	CR566920 CR566920
19	65.6	4.2	350	8	AQ474914 CTTR1-EI-
20	65	4.2	363	1	AV591350 AV591350
21	65	4.2	610	2	BE590137 197111.BA
22	64.2	4.2	587	7	CN322716 AGENCOURT
23	63.8	4.1	1203	9	CNS015Y4 Drosophi1
24	63.6	4.1	807	7	CK704772 ZP101-P00

25	63.6	4.1	888	9	AG030591 Pan trogl
26	63.4	4.1	783	5	BX915120 BX915120
27	63.2	4.1	406	1	AA792266 vn94e11.x
28	63.2	4.1	485	6	CF161888 B0703B06-
29	63.2	4.1	514	6	CA893864 B0181H02-
30	63.2	4.1	562	6	CA564994 K0332G02-
31	63.2	4.1	564	2	BB654405 BB654405
32	63.2	4.1	604	6	CD551263 B0325H01-
33	63.2	4.1	625	7	BU705998 UI-M-F00-
34	63.2	4.1	635	5	BU055310 UI-M-F00-
35	63.2	4.1	723	6	CB246896 UI-M-F00-
36	63.2	4.1	730	7	CF726532 UI-M-F00-
37	63.2	4.1	736	7	CN534827 UI-M-HS0-
38	63.2	4.1	754	7	CO045219 UI-M-HO0-
39	63.2	4.1	757	4	CN536888 UI-M-HO0-
40	63.2	4.1	762	6	CA327845 UI-M-FY0-
41	63.2	4.1	768	6	CA749488 UI-M-FY0-
42	63.2	4.1	787	7	CK637900 UI-M-HO0-
43	63.2	4.1	799	6	CA752497 UI-M-F00-
44	63.2	4.1	832	5	BQ771341 UI-M-F10-
45	63.2	4.1	855	5	BU152563 AGENCOURT
46	63.2	4.1	935	9	CNS0065XK Drosophi1
47	63.2	4.1	1066	3	CR715927 Tetradon
48	63.2	4.1	3137	3	AK034549 Mus muscu
49	63.2	4.1	925	9	CNS0091P Drosophi1
50	62.8	4.1	587	4	BG511885 dadd29d08.
51	62.6	4.1	677	7	CN526117 UI-M-HN0-
52	62.2	4.0	680	6	CB523529 UI-M-GH0-
53	62.2	4.0	751	5	BU056500 UI-M-F00-
54	62.2	4.0	776	9	AG561783 Mus muscu
55	61.8	4.0	262	9	CE173623 tigr-gss
56	61.6	4.0	434	2	BE234617 141780.MA
57	61.4	4.0	670	7	CF792818 885797.MA
58	61.4	4.0	672	7	CF794185 889669.MA
59	61.4	4.0	826	7	CN153741 940889.MA
60	61.4	3.9	520	4	BG307509 f159810.x
61	61	3.9	429	5	BY389139 BY389139
62	60.8	3.9	804	7	CNS35140 UI-M-HS0-
63	60.8	3.9	935	9	CNS06051 Drosophi1
64	60.4	3.9	359	2	BE814801 OV4-BN009
65	60	3.9	461	6	CF137810 UI-HF-BN0
66	60	3.9	544	5	BU429642 UI-HF-BN0
67	60	3.9	593	7	CN483764 h37c03.y
68	60	3.9	593	7	BE259148 B01108075
69	60	3.9	683	1	AU127678 AU127678
70	60	3.9	709	4	BG329665 602429276
71	60	3.9	731	2	BE258675 601116344
72	60	3.9	750	2	BE176589 RC3-HT058
73	60	3.9	752	2	BE297517 601178063
74	60	3.9	817	2	BE259148 B01108075
75	60	3.9	878	5	BQ437443 AGENCOURT
76	60	3.9	948	2	BE299945 600944490
77	60	3.9	974	4	BG394104 602456272
78	60	3.9	989	4	BM457082 AGENCOURT
79	60	3.9	1011	2	BF309509 601892016
80	60	3.9	1165	2	BQ213031 AGENCOURT
81	59.8	3.9	426	5	BF847408 PM4-EN006
82	59.8	3.9	885	5	BU193516 AGENCOURT
83	59.8	3.9	918	5	BU527069 AGENCOURT
84	59.8	3.9	943	5	BU228543 AGENCOURT
85	59.4	3.8	1104	5	BU722521 AGENCOURT
86	58.8	3.8	1132	5	BQ672163 AGENCOURT
87	58.4	3.8	1350	4	BM450237 AGENCOURT
88	58.4	3.8	416	6	CB606244 AMGNNUC:S
89	58.4	3.8	932	5	CN600720 Drosophi1
90	58.2	3.8	302	9	BP761124 BP761124
91	58.2	3.8	357	7	CF535776 UI-M-GH0-
92	58.2	3.8	361	7	CF535776 UI-M-GH0-
93	58.2	3.8	428	1	AA839768 vW51d11.x
94	58.2	3.8	452	6	CB520637 UI-M-G10-
95	58.2	3.8	575	7	CF911441 A0615D04-
96	58.2	3.8	690	5	BU703948 UI-M-F00-
97	58.2	3.8	2253	3	AK045449 Mus muscu

98	58	3.8	932	9	CNS00720	AL066742 Drosophila
99	58	3.7	997	5	AG057392	AG057392 Pan trogl
100	57.8	3.7	1038	5	BQ953925	BQ953925 AGENCOURT
C 101	57.6	3.7	1063	9	CL498639	CL498639 SAHL_657
102	57.4	3.7	745	7	CK693917	CK693917 ZP101-P00
103	57.2	3.7	251	5	BU056338	BU056338 UI-M-F00-
104	57.2	3.7	494	7	CF537874	CF537874 UI-M-G10-
105	57.2	3.7	1094	7	CK161715	CK161715 PGAS01428
106	57	3.7	993	2	BE874434	BE874434 60148870
C 107	56.8	3.7	1148	9	AG346700	AG346700 Mus muscu
C 108	56.8	3.7	1359	9	CL507830	CL507830 SAHL_788
109	56.6	3.7	459	6	CF143165	CF143165 UI-HF-BR0
110	56.6	3.7	540	2	AM502232	AM502232 UI-HF-BR0
111	56.6	3.7	862	4	BG758185	BG758185 602712357
112	56.6	3.7	934	5	BK410043	BK410043 BX410043
C 113	56.4	3.7	1523	9	AG448267	AG448267 Mus muscu
114	56.2	3.6	208	6	CF182368	CF182368 UI-M-EY0-
115	56.2	3.6	400	1	AJ683050	AJ683050 AJ683050
116	56.2	3.6	579	7	CN482597	CN482597 hmw2c09.Y
117	56.2	3.6	740	7	CF539286	CF539286 UI-M-G10-
118	56.2	3.6	798	6	CD352609	CD352609 UI-M-G10-
119	56.2	3.6	820	1	AU118164	AU118164 AU118164
C 120	56	3.6	665	9	AG159109	AG159109 Pan trogl
C 121	56	3.6	1355	9	AG429718	AG429718 Mus muscu
122	55.8	3.6	452	2	BF392580	BF392580 UI-R-CA0-
123	55.8	3.6	535	2	BF416810	BF416810 UI-R-CA0-
124	55.8	3.6	537	2	BF412941	BF412941 UI-R-CA0-
125	55.8	3.6	547	2	BE115719	BE115719 UI-R-CA0-
126	55.8	3.6	550	2	BF394139	BF394139 UI-R-CA0-
127	55.8	3.6	554	2	BF418125	BF418125 UI-R-CA0-
128	55.8	3.6	567	2	BE107720	BE107720 UI-R-CA0-
129	55.8	3.6	570	2	BF416261	BF416261 UI-R-CA0-
C 130	55.8	3.6	1462	9	AG441877	AG441877 Mus muscu
131	55.6	3.6	540	2	BF899692	BF899692 QV1-MT022
132	55.6	3.6	957	9	AG077010	AG077010 Pan trogl
133	55.6	3.6	1070	6	CA987146	CA987146 AGENCOURT
134	55.6	3.6	1312	4	BM914111	BM914111 AGENCOURT
135	55.4	3.6	1853	9	AG382854	AG382854 Mus muscu
C 136	55.4	3.6	1890	9	AG346134	AG346134 Mus muscu
137	55.2	3.6	392	6	CA876606	CA876606 K094P05-
138	55.2	3.6	426	6	CA871588	CA871588 K091LB08-
139	55.2	3.6	537	1	AI036739	AI036739 uc90a11.Y
140	55.2	3.6	566	7	CN688217	CN688217 E0254G07-
141	55.2	3.6	646	2	BB585201	BB585201 BB585201
C 142	55.2	3.6	1203	9	CNS01574	AL106054 Drosophila
143	55.2	3.6	1462	9	AG441877	AG441877 Mus muscu
144	55	3.6	412	2	BE018328	BE018328 b578903.Y
145	55	3.6	872	9	AG435681	AG435681 CGVGR23TV
C 146	55	3.6	1961	9	AG435522	AG435522 Mus muscu
147	54.8	3.5	485	9	CE447963	CE447963 t19r-988-
148	54.8	3.5	485	9	CE662468	CE662468 t19r-988-
149	54.8	3.5	899	9	CC700175	CC700175 OGLBQ18TH
C 150	54.8	3.5	965	8	AQ893417	AQ893417 HS_4832.A
151	54.8	3.5	1089	5	BQ648572	BQ648572 AGENCOURT
152	54.8	3.5	1280	9	CL470317	CL470317 SAHL_140
153	54.8	3.5	1552	9	AG430101	AG430101 Mus muscu
154	54.6	3.5	420	1	AL120275	AL120275 DKFP761L
155	54.6	3.5	457	2	BB852481	BB852481 BB852481
156	54.6	3.5	502	2	BF286221	BF286221 EST450812
157	54.6	3.5	531	2	BF899690	BF899690 QV1-MT022
C 158	54.6	3.5	583	4	BG375097	BG375097 UI-R-CV1-
159	54.6	3.5	645	4	BI380872	BI380872 BFLG1_002
160	54.6	3.5	720	7	CO400030	CO400030 AGENCOURT
161	54.6	3.5	842	9	AG058791	AG058791 Pan trogl
C 162	54.6	3.5	1101	9	CNS01783	AL108460 Drosophila
C 163	54.6	3.5	1628	9	CG757066	CG757066 P052-2-A0
C 164	54.4	3.5	519	2	BF899766	BF899766 QV1-MT022
C 165	54.4	3.5	1112	6	CA477644	CA477644 AGENCOURT
166	54.2	3.5	284	5	BQ436154	BQ436154 AGENCOURT
C 167	54.2	3.5	301	2	BF800390	BF800390 IL5-CI000
C 168	54.2	3.5	308	2	AM513912	AM513912 x090d06.x
C 169	54.2	3.5	342	7	T05980	T05980 EST03865 Fe
170	54.2	3.5	529	7	CN410758	CN410758 170005315

171	54.2	3.5	618	4	BG574441	BG574441 602596388
172	54.2	3.5	682	2	AM961105	AM961105 EST373072
173	54.2	3.5	689	1	AV707144	AV707144 AV707144
174	54.2	3.5	701	4	BM788417	BM788417 R-EST0067
175	54.2	3.5	822	5	BU568076	BU568076 AGENCOURT
176	54.2	3.5	824	4	BG565128	BG565128 602583752
177	54.2	3.5	870	4	BG614027	BG614027 602641565
C 178	54	3.5	1569	9	AG341503	AG341503 Mus muscu
C 179	53.8	3.5	742	9	CG696734	CG696734 MUGO_CH25
180	53.8	3.5	809	4	BI259307	BI259307 602972544
181	53.8	3.5	1265	9	CG750734	CG750734 P045-2-B0
182	53.8	3.5	1314	9	AG039591	AG039591 Pan trogl
183	53.6	3.5	249	5	BQ308190	BQ308190 MR0-BT450
C 184	53.6	3.5	611	4	BI044950	BI044950 MR4-OT010
C 185	53.6	3.5	662	7	CN277795	CN277795 170005313
C 186	53.6	3.5	726	9	BM174921	BM174921 Dantio ter
187	53.6	3.5	853	5	BUB55947	BUB55947 AGENCOURT
188	53.6	3.5	924	4	BG395346	BG395346 602457885
189	53.6	3.5	1004	5	BU149386	BU149386 AGENCOURT
C 190	53.6	3.5	1137	4	BG809979	BG809979 mgc002xd
C 191	53.6	3.5	1830	9	CL078620	CL078620 CH216-151
C 192	53.6	3.5	2243	9	AG381986	AG381986 Mus muscu
C 193	53.4	3.5	428	5	BX916095	BX916095 BX916095
C 194	53.4	3.5	428	5	BX922861	BX922861 BX922861
195	53.4	3.5	949	9	AG171092	AG171092 Pan trogl
196	53.4	3.5	1452	9	AG032979	AG032979 Pan trogl
197	53	3.4	738	5	BU326843	BU326843 603492618
198	53	3.4	1088	4	BM450730	BM450730 AGENCOURT
199	53	3.4	1093	4	BM803558	BM803558 AGENCOURT
C 200	53	3.4	1160	9	AG043473	AG043473 Pan trogl

ALIGNMENTS

RESULT 1
BX112453
LOCUS
BX112453 NCI CGAP GC4 Homo sapiens cDNA clone IMAGE998F231165 ;
DEFINITION
IMAGE:1257620, mRNA sequence.

ACCESSION
BX112453
VERSION
BX112453.1 GI:27837617
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 642)
Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,
Radelof, U., Schneider, D. and Korn, B.

TITLE
Human Unigeneset - RZPD3
JOURNAL
Unpublished (2003)
COMMENT
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGE998F231165.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi?responsefileNo=972
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
Location/Qualifiers
1..642
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

FEATURES

source

ORIGIN

```

/clone="IMAGp988F213165 ; IMAGE:1257620"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_1ib="NCI CGAP GC4"
/polymer="Vector: pTR73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pTR73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo. "

Query Match      39.6%; Score 612.4; DB 5; Length 642;
Best Local Similarity 99.7%; Pred. No. 1.3e-127;
Matches 624; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 916 GTGAGAGAACTGGCTCTGTGGCGGTGCGGGGTGCGGAGATGATGCGAGCTGC 975
DB 1 GTGAGAGAACTGGCTCTGTGGCGGTGCGGGGTGCGGAGATGATGCGAGCTGC 60
QY 976 TGGGTGTACTCACTGCGCGCTGCTTCCACTGCGCTGCCACTTCCAGCCGCACT 1035
DB 61 TGGGTGTACTCACTGCGCGCTGCTTCCACTGCGCTGCCACTTCCAGCCGCACT 120
QY 1036 CCCGCGCCGGGACGGGCGCTGCGCTGCAAGATCTCTCAGAGAGAGTGAAGCCGCTG 1095
DB 121 CCCGCGCCGGGACGGGCGCTGCGCTGCAAGATCTCTCAGAGAGAGTGAAGCCGCTG 180
QY 1096 TGGAG-GGGGGTGTGGCGCCCAAGCCCGCGCTGCGGCGCTGCGGCGCTGCAAGATGAC 1154
DB 181 TGGAGGGGGTGTGGCGCCCAAGCCCGCGCTGCGGCGCTGCGGCGCTGCAAGATGAC 240
QY 1155 ACTGCAAGTACAGAGCCGCTGTCACAGAGGATGACCTGAGTCCCTTTCAGAGGAC 1214
DB 241 ACTGCAAGTACAGAGCCGCTGTCACAGAGGATGACCTGAGTCCCTTTCAGAGGAC 300
QY 1215 ACCTTGATGAGCATCTCTGAGTGGGCCATCAGAGCATGGCCCGTCCGCGGCGCTTTC 1274
DB 301 ACCTTGATGAGCATCTCTGAGTGGGCCATCAGAGCATGGCCCGTCCGCGGCGCTTTC 360
QY 1275 CCTCTCTGACCCCAAGATGGCGCGGAGCATGAGCTGTGAGAGAGTGTGAGAGAGCA 1334
DB 361 CCTCTCTGACCCCAAGATGGCGCGGAGCATGAGCTGTGAGAGAGTGTGAGAGAGCA 420
QY 1335 CCTCTCTCTCACTCTGAGAGCGGCGGCTGAGATCAAGAGAGGAGCAAGCCCACTC 1394
DB 421 CCTCTCTCTCACTCTGAGAGCGGCGGCTGAGATCAAGAGAGGAGCAAGCCCACTC 480
QY 1395 TTGTCAAGTGTGGCTGTAAAGCTCTGTGTTTGTGGGAGACAGCCATCATGTGCTT 1454
DB 481 TTGTCAAGTGTGGCTGTAAAGCTCTGTGTTTGTGGGAGACAGCCATCATGTGCTT 540
QY 1455 GGAATTTAAACCTGGCCCACTTCTACTCTGAAAGTCCCGGAGGCTCTCTTGCTT 1514
DB 541 GGAATTTAAACCTGGCCCACTTCTACTCTGAAAGTCCCGGAGGCTCTCTTGCTT 600
QY 1515 GGTGACCTACTAAATATATTAAT 1540
DB 601 GGTGACCTACTAAATATATTAAT 626

RESULT 2
LOCUS AY419550 1506 bp DNA linear GSS 17-DEC-2003
DEFINITION Homo sapiens AIRE gene, VIRUTAL TRANSCRIPT, partial sequence,
ACCESSION AY419550
VERSION AY419550.1 GI:39775507
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

ORIGIN

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 1506)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Fertier,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL
Science 302 (5652), 1960-1963 (2003)
PUBMED
14671302
REPERENCE
AUTHORS
2 (bases 1 to 1506)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Fertier,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Direct Submision
JOURNAL
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
location/Qualifiers
source
1..1506
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1506
/gene="AIRE"
/locus_tag="HCM6928"

Query Match      35.1%; Score 542.6; DB 9; Length 1506;
Best Local Similarity 71.0%; Pred. No. 8.5e-112;
Matches 545; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

QY 516 CTGTGCGAAGAAATGAGACGAGTGTGCGGTGTGCGGAGCGGAGCTCATCTGC 575
DB 739 CTTCACCAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 798
QY 576 TGTGACGGTGTGCGCTTCCACTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 635
DB 799 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 858
QY 636 CCCAGTGGACCTGAGAGTGTCTCAGCTGCTGCAAGCAAGTCCAGAGTGTGAGCC 695
DB 859 NNNNNNTGGACCTGAGAGTGTCTCAGCTGCTGCAAGCAAGTCCAGAGTGTGAGCC 918
QY 696 CGGAGAGAGAGCCCGGCGCCAGAGAGCACCCTGAGAGACCCCGCTCCCGGAGCTT 755
DB 919 CGGAGAGAGAGCCCGGCGCCAGAGAGCACCCTGAGAGACCCCGCTCCCGGAGCTT 978
QY 756 AGGTGGCGGAGAGAGAGTGAAGTCCACTGGGGAACCCCTAGCCGAGATGACAG 815
DB 979 AGGTGGCGGAGAGAGAGTGAAGTCCACTGGGGAACCCCTAGCCGAGATGACAG 1038
QY 816 ACTCTTGTCTAAGACACTGCGGCTCGCTTCTGAGAGCCCGCTGCAAGTGTGAG 875
DB 1039 ACTCTTGTCTAAGACACTGCGGCTCGCTTCTGAGAGCCCGCTGCAAGTGTGAG 1098
QY 876 TCTGGGCGCTGACACCCCTACTGTTGTTGGGTCCCTGAGAGGTGAGAGAACTGCTCT 935
DB 1099 TCTGGGCGCTGACACCCCTACTGTTGTTGGGTCCCTGAGAGGTGAGAGAACTGCTCT 1158
QY 936 GGTGCGCTTGTGCGGGTGTGCGAGATGATGAGAGTGTGCTGCTGCTGCTGCTGCT 995
DB 1159 GGTGCGCTTGTGCGGGTGTGCGAGATGATGAGAGTGTGCTGCTGCTGCTGCTGCT 1218
QY 996 GCTGCTTTCACATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
DB 1219 GCTGCTTTCACATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1278
QY 1056 CGGTGAGATCTGTGCTCAAGAGAGTGAAGCCAGCCCTGTGTGAGAGGGGTGTGCGCC 1115

```

[illegible][illegible]

Query Match	5.6%	Score 86	DB 1	Length 384	
Best Local Similarity	77.6%	Pred. No. 1.1e-09			
Matches	104	Conservative	0	Mismatches	30
				Indels	0
				Gaps	0
Oy	1148	GGATGACACTGCCAGTCAAGACCCGCTCTTGACAGGAGTACCTGAGTCCCTTGAG	1207		
Db	372	GGACGACTCTGCTAGTCAAGACCCCTGTTCTACATAGGAGACACCTGGAGTCCCTCTCA	313		
Oy	1208	CGAGCACACTTCCATGGCATCTCTGCACTGGGACCATCCAGAGCATATGGCCGCTCCGGGCGC	1267		
Db	312	TGAGCACTCATTTGACGGCATCTGCACTGGGACCATCCAGAGCATATGACGCCGCTGAC	253		
Oy	1268	CCCTTCCCTCCT 1281			
Db	252	CGAGACACACCTCT 239			
RESULT 13					
AG069825	947 bp	DNA	linear	GSS 03-NOV-2001	
LOCUS	Pan troglodytes DNA, clone: PTB-060113.F, genomic survey sequence.				
ACCESSION	AG069825				
VERSION	AG069825.1	GI:16621627			
KEYWORDS	GSS.				
SOURCE	Pan troglodytes (chimpanzee)				
ORGANISM	Pan troglodytes				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.				
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,				
JOURNAL	Totoki, Y., Watanabe, H. and Sakaki, Y.				
REFERENCE	BAC end sequences of library PTB				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 947)				
JOURNAL	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,				
REFERENCE	Totoki, Y., Watanabe, H. and Sakaki, Y.				
AUTHORS	Direct Submission				
TITLE	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical				
JOURNAL	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);				
REFERENCE	1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan				
AUTHORS	(E-mail: chimdee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,				
TITLE	Tel: 81-45-503-9111, Fax: 81-45-503-9170)				
JOURNAL	Clones are derived from the chimpanzee BAC library PTB. This BAC end				
REFERENCE	was generated during the RAD process and may have higher chance of				
AUTHORS	clone tracking errors.				
TITLE	PRIMERS				
JOURNAL	Sequencing: -21M13				
REFERENCE	LIBRARY				
AUTHORS	Vector : pKS145				
TITLE	R.Site 1 : SacI				
JOURNAL	R.Site 2 : SacI.				
REFERENCE	Location/Qualifiers				
AUTHORS	1. 947				
TITLE	/organism="Pan troglodytes"				
JOURNAL	/mol_type="genomic DNA"				
REFERENCE	/db_xref="taxon:9598"				
AUTHORS	/clone="PTB-060113.F"				
TITLE	/sex="male"				
JOURNAL	/cell_type="lymphoblast"				
REFERENCE	/clone_lib="PTB Chimpanzee Male BAC library"				
ORIGIN					
Query Match	4.8%	Score 73.4	DB 9	Length 947	
Best Local Similarity	43.0%	Pred. No. 8.5e-06			
Matches	338	Conservative	0	Mismatches	448
				Indels	0
				Gaps	0
Oy	606	GCTGCTGTGCTCCCTCGCTCGCGAGATCCCACTGGAGCTGAGGTGCTCCAGCTGC	665		
Db	136	GAGCGCCCGGAGCGCGCGCAGNCCCGGCGCCCCCGGAGCGGAGGAGGAGGCGGAGG	195		
Oy	666	CTGACGAGCAAGTCCAGAGGTGAGGCCCCGGGAGAGAGAGAGCCCGGCCCCAGAGCA	725		

D	b		196	GAGGAGCGCCGGAGCGCGGCAGGCCGCCGCCCGCGCGGAGGAGGAGGAGCGCGCGCCCG	255
Oy			726	CCCGTGGAGAACCCTGCTCCCCCGGGGTCTTAGTGTCGGCGGGAGAGAAGAGGTAAAGGTCCA	785
D	b		256	GCGCGCGGGGGCGAAGCGCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	315
Oy			786	CTTGGAGAACCCCTAGCGCGCATATGAACACGACTTTTGTATCAAGACAATACTGCGGACTCCG	845
D	b		316	GGGGGGGGCCCGCGCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	375
Oy			846	CCTTCTGACAGCCCGCTGCGCAGGTCTGAAATCTCTCGGCTTGACACCCCCTTAATGTGTGTG	905
D	b		376	GCG	435
Oy			906	GGTCTTAGAGGGTACAGCAACAATTGAGTCTCTGAGTGGCGGTTGGCGGGGTGTGCGAGATAGT	965
D	b		436	GCCCCCG	495
Oy			966	ACGACGCTGTGCGGNTACTCACTGCGCGCTTCTCACTAGCGGCTCCCACTTCCCA	1025
D	b		496	AGGCG	555
Oy			1026	GCCCGCACCTTCCCG	1085
D	b		556	GCGCGGGGGCG	615
Oy			1086	CNAACCCCTGTGAGAGGGAGTGTGACCCGCCAGCCGCGCGCGCGCGCGCGCGCGCGCTGCC	1145
D	b		616	CGGCG	675
Oy			1146	AAGATATACACTGCGCAATCACAGACCCCGCTCTGTGCACAGGATATGACTGAGTCCCTTCTG	1205
D	b		676	GCCCG	735
Oy			1206	AGCAGACACACTTTCATGTGCATCTTGCATGTGGGCGATCCAGAGATAGCGCTCCGGCG	1265
D	b		736	CCGAGAGCG	795
Oy			1266	GCCCCCTTCCCCCTCTTAACCCCAATGCGCGGGACATGACCTTATATGAGAAGATGCTG	1325
D	b		796	GCCCCCG	855
Oy			1326	AGAAAGACACCTCTTCTCTAGTCTCTGGAAGCGCGCGCTGGGATCAAGAAGGGAGCAG	1385
D	b		856	GNCG	915
Oy			1386	CGCCAC	1391
D	b		916	GGCGCG	921
<hr/>					
RESULT 14					
CNS03FKO			959 bp	DNA	linear GSS 01-SEP-2000
LOCUS			Tetradon nigriviridis genome survey sequence PUC-Ori end of clone		
DEFINITION			022016 of library G from Tetradon nigriviridis, genomic survey		
			sequence.		
ACCESSION			AL241809		
VERSION			AL241809.1		GI:7962578
KEYWORDS			GSS; genome survey sequence.		
SOURCE			Tetradon nigriviridis		
ORGANISM			Tetradon nigriviridis		
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
			Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
			Acanthomorpha; Acanthopterygii; Percomorphia; Tetraodontiformes;		
			Tetraodontidae; Tetraodontidae; Tetraodon.		
REFERENCE			1		
AUTHORS			Roeft Collinus,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,		
			Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,		
			Saurin,W. and Weissenbach,J.		
TITLE			Estimate of human gene number provided by genome-wide analysis		
JOURNAL			using Tetradon nigriviridis DNA sequence		
			Nat. Genet. 25 (2), 235-238 (2000)		

MEDLINE	20296633
PUBMED	10835645
REFERENCE	2
AUTHORS	Roebe Crollius,H., Jallion,O., Dabily,C., Ozouf-Costaz,C., Plameres,C., Fischer,C., Bouneau,T., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Genome Res. 10 (7), 939-949 (2000)
TITLE	
JOURNAL	
MEDLINE	20359837
PUBMED	10899143
REFERENCE	3 (bases 1 to 959)
AUTHORS	Genoscope.
TITLE	Direct Submision
JOURNAL	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .
COMMENT	
FEATURES	
SOURCE	location/Qualifiers 1..959 /organism="Tetraodon nigroviridis" /mol_type="genomic DNA" /db_xref="taxon:99883" /clone="022016" /clone_1lb="G" /note="Genoscope sequence ID : COBG022BH08SP1-end : PUC-Ori"
ORIGIN	
Query Match	4.7% ; Score 73 ; DB 9 ; Length 959 ;
Best Local Similarity	60.8% ; Pred. No. 1.1e-05 ;
Matches 110 ;	Conservative 5 ; Mismatches 66 ; Indels 0 ; Gaps 0 ;
Oy	552 CAGAAAGATGAGACGAGATGTGCCCTGTGTCCGGACCGCGGGAAGCTCATCTGCTGTAC 581
Db	434 CTGGAGCATTGAGGACCACTGTGCCTCTGCGAGGAGACGAGACTCAGCCCTGCCGC 493
Oy	582 GGCTTCCTCCGGGACCTTCACCTGACCTGCTGCTGCTCCCTCCGCTCCGGAAGATCCCAGT 641
Db	494 AGCTGCTVCCCGGCTTTMACCCCCAGCTGTCTGCACCCCCCAMTAATAAAGCCCCCGG 553
Oy	642 GGGACCTGAGGTGCTCCAGCTGCTCTGCAAGGACAAGTCCAGAAGTGTCAGAGCCCGGACA 701
Db	554 GGCCCTGTGTACTGCCCAAGTGCCAGAAAGTAGTACTTCTGCTGCCACGTAAGTCCA 613
Oy	702 G 702
Db	614 G 614
RESULT IS	
LOCUS	CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T813 end of BAC #
KEYWORDS	BACH19D16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
VERSION	AL053013
KEYWORDS	AL053013.1 GI:4934461
ORGANISM	GSS.
REFERENCE	Drosophila melanogaster (fruit fly)
AUTHORS	Drosophila melanogaster
TITLE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pezomyza;
JOURNAL	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyndroidea; Drosophilidae; Drosophila. 1 (bases 1 to 925)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : :
JOURNAL	BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT
Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library](http://www.fruitfly.org/The%20BDGP%20Drosophila%20melanogaster%20BAC%20library%20was%20prepared%20by%20Kazuhiro%20Osoegawa%20and%20Aaron%20Mammoser%20in%20Piet%20de%20Jong's%20laboratory%20in%20the%20Department%20of%20Cancer%20Genetics%20at%20the%20Roswell%20Park%20Cancer%20Institute%20in%20Buffalo,%20NY) was prepared by Kazuhiro Osoegawa and Aaron Mammoser in Piet de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp. The same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 Source
 Location/Qualifiers
 1..925
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR19D1c"
 /clone_1fb="RPCI-98"
 /note="End : TET3"

ORIGIN
Query Match 4.5%; Score 69.6; DB 9; Length 925;
Best Local Similarity 13.5%; Pred. No. 6.2e-05;
Matches 46; Conservative 173; Mismatches 123; Indels 0; Gaps 0;

OY 833 CTTCCGCGCTCCGCTTCTGCAGCCCGCGCTGCAGATCTGCAGCTTCGCGCTGCAGACC 892
 575 CSCGSCSSSSGSCBCCCCGSSSYCCSSBSBSKSTBSGSCCCSSKSVCTGSSSS 634
OY 893 CCTACTGTGTGTGGTCTCTGAGGCTCAGCAAACTGCTCTGTGCGCGCTTGCGGGGT 952
DB 635 SSCGSSSSSSSTSSSTSSSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 694
OY 953 GTGCGGAGATGTACGACGACTGCTGCGGTACTACTGCGCGCGCTTCCACTGAGC 1012
DB 695 TSSSSSSSSSTSSSSSVSGSKSTBSBSGSSSSSSSSSSSSSSSSSSSSSSSSSSSS 754
OY 1013 CTGCGCACTTCCGACCGGACCTCCGCGCGGACGCGGCGCTGCGTGCAGATCTGCTC 1072
DB 755 SSTGSCCTCCSSSYSSSTSSSSSTSSGSSSSSVGTRSSSDSTSTCCSCCTWCTC 814
OY 1073 AGGAGACGTGACCCAGACCCCTGTGAGAGGGGTGCTGACCCCGCGCGCGCTGAC 1132
DB 815 CSTYMBGCTSTSCGSSSSSGKGKVTKCGCGCGGSSSTNGMBGTSTACSSSSSSSSSS 874
OY 1133 CCCTGGGCGCTGCCAAGATGACACTGCCGATGCACGACCCGC 1174
DB 875 VSSSSKSSASSSVSSSGSGVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 916

RESULT 16
BU740688/c

LOCUS BU740688 674 bp linear EST 10-OCT-2002
DEFINITION UI-E-EJ0-ai1-m-23-0-UI.a1 UI-E-EJ0 Homo sapiens cDNA clone
 BU740688 1 (bases 1 to 674)
ACCESSION BU740688
VERSION BU740688.1 GI:23680847
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 674)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINS 97044477
PUBMED 8889548

Df		97	CCTGTGCGCAGCACCTCCGCCAGGGGGCTTACATCATCTCGTGNGCCCTGGACCAGAGCTGGAGA	156
Oy		632	GATCCCCCACTGGAGACTTGAGTGTCTTCAGCTGCCTCTGCAGGACAACAGTTCAGAGAGTGA	691
Df		157	GCGTCCCAGGAGCAAGTGAAGCTGCCCCCAGCTGTGAAGAGGGGAAATTCAGTGGAGCC	216
Oy		692	G	692
Df		217	G	217
RESULT 18				
LOCUS		CRS66920	940 bp	mRNA EST 19-JUL-2004
DEFINITION		CRS66920	XGC-tailbud-head Xenopus tropicalis cDNA clone THdA010e06	
ACCESSION		CRS66920		
VERSION		CRS66920.1	GI:50396997	
KEYWORDS		EST.		
SOURCE		Xenopus tropicalis (western clawed frog)		
ORGANISM		Xenopus tropicalis		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;		
		Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoes; Pipidae;		
		Xenopodinae; Xenopus; Silurana.		
REFERENCE		1 (bases 1 to 940)		
AUTHORS		Croking,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.		
JOURNAL		Sanger Xenopus tropicalis EST project 2001 (2004)		
COMMENT		Unpublished (2004)		
		Contact: Croking MDR		
		Sanger Institute		
		Hinxton, Cambridgeshire, CB10 1SA, UK		
		Email: trop@sanger.ac.uk		
		Sanger Xenopus tropicalis EST project 2001		
		TROPICALIS SEQUENCE ID: THdA010e06.plkSP6		
		This sequence is from a Xenopus Gene Collection (XGC) library		
		constructed by Nigel Garrett.		
		Seq primer: SP6.		
FEATURES				
source		Location/Qualifiers		
		1..940		
		/organism="Xenopus tropicalis"		
		/mol_type="mRNA"		
		/db_xref="taxon:8364"		
		/clone="THdA010e06"		
		/dev_stage="tailbud head (stage 28-30)"		
		/lab_host="Escherichia coli DH10B."		
		/clone_1lb="XGC-tailbud-head"		
		/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA		
		was oligo dT primed from 5ug of poly A+ RNA from tailbud		
		head. EcoRI-NotI cut cDNA was then ligated into pCS107		
		with EcoRI at the 5' end and NotI at the 3' end."		
ORIGIN				
Query Match		4.3%; Score 67; DB 7; Length 940;		
Best Local Similarity		66.0%; Pred. No. 0.00024;		
Matches		97; Conservative 0; Mismatches 50; Indels 0; Gaps 0,		
Oy		519	TGCCAAGAATGAGAGCAGGTGTGTCCTGTGTCCGACGCGGGAGACTATCTGCTGT	578
Df		245	TACCAAGATTAAAGACGATGAGTGTCTCAGTGTCCAAGATGCGGGGAGATTATATGTTGCC	304
Oy		579	GACGAGTCCCTCGGGGCTTCACCTGTGGCTGCTGTCCCTCCGCTCCGGAGAGATCCCC	638
Df		305	GATGATGCCCAACGGTCTTTACCTTTCTGCTGTGTCGCGCTTTTACCATATTTCCA	364
Oy		639	AGTGGACCTGGAGGTCTCCACGCTGC	665
Df		365	AGCGGCACATGAGATGTGATCTTGC	391
RESULT 19				
LOCUS		A0474914	350 bp	DNA linear GSS 23-APR-1995
DEFINITION		CITBI-51-2591L22.TF CITBI-51 Homo sapiens genomic clone 2591L22.		

ACCESSION	genomic survey sequence.	
VERSION	A0474914	
KEYWORDS	A0474914.1	GI:4657033
SOURCE	GSS.	
ORGANISM	Homo sapiens (human)	
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
AUTHORS	1 (bases 1 to 350) Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.	
TITLE	Use of BAC End Sequences from CalTech Libraries for Sequence-Ready Map Building Unpublished (1997)	
JOURNAL	Other GSSs: CITR1-RI-2591L22.TR Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208	
COMMENT	Email: hped@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html . Seq primer: M13-21 Class: BAC ends.	
FEATURES	Location/Qualifiers	
SOURCE	1..350 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="2591L22" /sex="male" /cell_type="sperm" /clone_lib="CITR1-RI" /note="Vector: pBelBAC11; Site_1: EcoRI; Site_2: EcoRI; Caltech Human BAC Library D"	
ORIGIN		
Query Match	4.2%; Score 65.6; DB 8; Length 350;	
Best Local Similarity	94.4%; Pred. No. 0.00044;	
Matches	68; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
OY	1145	CAGAGATGACCTGCCAGTCAAGCCCGCTCTGCACAGGGATGACTGAGTCCCTTCT 1204
DB	109	CCAGAGATGACCTGCCAGTCAAGCCCGCTCTGCACAGGGATGACTGAGTCCCTTCT 168
OY	1205	GAGCGAGACAC 1216
DB	169	GAGCGAGGTAAAC 180
RESULT 20		
LOCUS	AV591350	563 bp mRNA linear EST 27-NOV-2001
DEFINITION	AV591350 Bos taurus brain fetus Bos taurus cDNA clone E1BR018A07	
ACCESSION	AV591350	
VERSION	AV591350.1	GI:9702343
KEYWORDS	EST.	
SOURCE	Bos taurus (cow)	
ORGANISM	Bos taurus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.	
AUTHORS	1 (bases 1 to 563) Takaue,A., Hitotsume,S., Itoh,R., Itohono,A., Suzuki,H., Aso,H. and Sugimoto,Y.	
TITLE	Reestablishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs	
JOURNAL	Nucleic Acids Res. 29 (22), E108 (2001)	

MEDLINE 21570554
PUBMED 11713328
COMMENT Contract: Yoshitaku Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
OdaKura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel.: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES
source Location/Qualifiers
1..563
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/collection="E1BR018A07"
/tissue_type="brain"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_id="Bos taurus brain fetus"
/note="Vector: pZL1; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"

ORIGIN

```
Query Match          4.2%   Score 65; DB 1; Length 563;
Beet Local Similarity 56.2%; Pred. No. 0.00064;
Matches 122; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
```

QY	533 GGACGAGTGTCCCGTGTCGGGAGGAGGGAGCTCATCTGCTGAACGCTGCCCTGC	592
DB	28 GGAGTATTGCCGCGGTGCAMAGAGGAGGGAGTGCTGTCTCGACGCTGCATCTC	87
QY	593 GAGCTTCACACTGAGCTGCTGCTCCCTCCGCTCCGGAGATGCCAGTGGAGCCTTGAG	652
DB	88 CTCCTACCACTCATCTGTCTGAACCCCCCGCTGCCGACATCCCACAAGGGATGGCT	147
QY	653 GTGCTTCAGTCTGCTTCGACGACAACTGCAGAGGTGCAGCCCGGAGAGAGCCCGC	712
DB	148 GTGTCCCGAATGACACTTGTCTGTGCTCAAAGCCGTTGTGCAGAAGATCTGCACGTGCG	207
QY	713 GCCCCAGGAGCACCCGTGTGAGAGACCCTCCCTCCCGC	749
DB	208 GTGGGAGAGCCACCCCTGTGGCATGCCACGCCCCGAG	244

RESULT 21
BE590137 mRNA linear EST 27-MAR-2003
LOCUS BE590137 610 bp
DEFINITION 197111 BARC SBOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE590137
VERSION BE590137.1 GI:9843176
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 610)
Sonstegard,T., Capuco,A.V., White,J., Van Tassel,C.P.,
Connor,E.E., Cho,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D.
and Quackenbush,J.
Analysis of bovine mammary gland EST and functional annotation of
the Bos taurus gene index
Mamm. Genome 13 (7), 373-379 (2002)

TITLE
JOURNAL MEDLINS PUBMED
CONTACT: Sonstegard TS
USA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel.: 301 504 8416
Fax: 301 504 8414
Email: tats@ipsl.barc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred v0.98004.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCAGTCAGCAGC
 Plate: 117 row: I column: 4
 Seq primer: ATTAGGTGACACTATG.

FEATURES
 SOURCE
 1..610
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /feature_type="pooled"
 /lab_host="DH108"
 /clone_id="BARC.5B0V"
 /note="Vector: pCMV SPOT6; Site_1: NotI; Site_2: SalI;
 library made from pooled mRNA isolated from mammary
 tissues at eight physiological, developmental, and disease
 states."

ORIGIN

Query Match	4.2%	Score 65;	DB 2;	Length 610;
Bee local Similarity	62.7%	Pred. No. 0.00064;	Mismatches 60;	Indels 0; Gaps 0;
Matches 101;	Conservative 0;	Mismatches 60;	Indels 0;	Gaps 0;

Oy 532 AGACGAGTGTGCGGTGTGCGGAGCGGAGCTCATCTGTCGACGCTGCCCTC 591
 |||||
 Db 431 AGATTACTGTGAGTGTGCGAGCAGGAGTGAGGATCATCTGTGCAACCTGCCGA 490
 |||||
 Oy 592 GGGCTTCACACTGCGCTGCTGTCCTCCGCTCCGGAGATCCCAAGTGGACCTGGA 651
 |||||
 Db 491 GGGCTTACACTGCTGTGCTGACCCAGAGCTGAGAAAGCTCCGAGGCGCAATGGA 550
 |||||
 Oy 652 GTGCTTCAGCTGCTGCTGAGCAACAGTCCAGAGAGTGCAG 692
 |||||
 Db 551 GCTGCCCCCACTGTGAGAAAGAGGGATCCAGTGGAGCGG 591
 |||||

RESULT 22 CN322716 587 bp mRNA linear EST 14-APR-2004
 LOCUS CN322716/c
 DEFINITION AGENCOURT 21860720 Xtlst10-30 Xenopus tropicalis cDNA clone
 IMAGE:7092432 3', mRNA sequence.
 ACCESSION CN322716
 VERSION CN322716.1 GI:46380352
 SOURCE EST.
 ORGANISM Xenopus tropicalis (western clawed frog)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 587)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics / NIH
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: gsabes-r@mail.nih.gov
 Tissue Procurement: Dr. Wei Wu / Prof. Christof NIEHR
 cDNA Library Preparation: Wei Wu, DKFZ
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LHAM14930 row: k column: 22
 High quality sequence stop: 587.
 Location/Qualifiers
 1..587
 /organism="Xenopus tropicalis"

FEATURES
 SOURCE

Db 695 GCGGCGCC 702

RESULT 26
LOCUS BX915120
DEFINITION 783 bp mRNA linear EST 07-MAY-2004
BX915120 Sus scrofa library (scan) Sus scrofa cDNA clone
scan0028d.a.12 5prim, mRNA sequence.

ACCESSION BX915120
VERSION BX915120.1 GI:41131899
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa (pig)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE 1 (bases 1 to 783)
Bonnet, R., Tosser-Klopp, G., Benne, P., Cabau, C., Villedier, S.,
Soares, M., Donald, F. and Haley, F.
A Pig Normalised Multi-Tissue cDNA Library
Unpublished (2003)
CONTACT: Tosser-Klopp G
Genetique Animale
Institut National de la Recherche Agronomique
Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
cedex, FRANCE
Tel: 33 (0) 5.61.28.51.14
Fax: 33 (0) 5.61.28.53.08
Email: tosser@toulouse.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at sigenauport@jouy.inra.fr to obtain the chromatogram of this
sequence.
Plate: 0028 row: 8 column: 12.
Location/Qualifiers
1. .783
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="scan0028d.a.12"
/tissue_type="mixed"
/dev_stage="from embryos to adults"
/note="Library: Sus scrofa library (scan)"
/note="Tissues: adipose tissue, brain, kidney, liver,
muscle, ovary, testis, heart, hypothalamus, pancreas,
skin, spleen, thymus, placenta, pituitary gland, seminal
vesicle, small intestine, uterus, adrenals, bulbourethral
gland, cerebral trunk, epididymis, female gonad,
gall-bladder, hippocampus, large intestine, male gonad,
melanocytes, stomach, udder"

ORIGIN

Query Match 4.1%; Score 63.4; DB 5; Length 783;
Best Local Similarity 61.7%; Pred. No. 0.0015;
Matches 100; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 528 AATGAGAGAGAGTGTGCGGAGCGGAGAGCTATGCTGTGACGCTGC 587
|||
Db 585 AATGAGAGAGTGTGCGGAGCGGAGAGCTATGCTGTGAGAAAGTG 644
|||
QY 588 CCTCGGCGCTTCCAGCTGCGCTGCTGCTCCCTCCGTCGAGAGATCCCACTGGAGAC 647
|||
Db 645 CCGAAGTATTCACCTGCTGTGTCAAGTCCGACCGCTGCAATTTCCGAGTGGAGAG 704
|||
QY 648 TGGAGGTGCTCCAGCTGCTGCAAGCAAGTCCAGAGGTG 689
|||
Db 705 TGGATTGACATTTCTGCGAGACTTATCCNAACAGAAAGTG 746
|||

RESULT 27
LOCUS AA792266
DEFINITION 406 bp mRNA linear EST 09-FEB-1998
vn94e11.r1 StrataGene mouse heart (#937316) Mus musculus cDNA clone
IMAGE:1039628 5', mRNA sequence.

ACCESSION AA792266

VERSION AA792266.1 GI:2855221
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 406)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
CONTACT: Marras M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28m3 rev1 ET from Amersham
High quality sequence stop: 403.
Location/Qualifiers
1. .406
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NIH Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:1039628"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse heart (#937316)"
/note="Organ: heart; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
adaptor sequence: 5' GATTCCGCGACAG 3' ~3' adaptor
sequence: 5' CTCGAGTTTATTTTATTTT 3' "

ORIGIN

Query Match 4.1%; Score 63.2; DB 1; Length 406;
Best Local Similarity 67.4%; Pred. No. 0.0016;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GAGAGAGTGTGCGGAGCGGAGAGCTATGCTGTGACGCTGCCCTCG 592
|||
Db 220 GGAATTCGTCTCGTGTGCAAGAGCGGAGAGCTCTGTGTCATGACATGCGCTTC 279
|||
QY 593 GCGCTTCCAGCTGCGCTGCTGCTCCCTCCGTCGAGAGATCCCACTGGAGAC 652
|||
Db 280 TTCTTACACATTCACATGCTGTAACCCCGCTGCGAGATCCCAACGCGGAATGCT 339
|||
QY 653 GTGCTCCAGCTG 664
|||
Db 340 CTGTCCCGCTG 351
|||

RESULT 28
LOCUS CF161888
DEFINITION 485 bp mRNA linear EST 25-JUL-2003
B0703B06-5 NIA Mouse Embryonic Germ Cell cDNA library (long) Mus
musculus cDNA clone NIA:B0703B06 IMAGE:30458165 5', mRNA sequence.

ACCESSION CF161888
VERSION CF161888.1 GI:33271437
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS
TITLE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 485)
Piao, Y., Ko, N.-T., Lim, M. K. and Ko, M. S. H.
Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method

JOURNAL
MEDLINE
PUBMED
COMMENT

Genome Res. 11 (9), 1553-1558 (2001)
21429098
11544199
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Caswell Drive, Suite 4060, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: B0703 row: E column: 06
Seq primer: M13 Reverse
High quality sequence stop: 485
POLYA-No.

FEATURES
source

1. 485
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaBST:B0703B06-5"
/db_xref="taxon:10090"
/clone="NIA:B0703B06 IMAGE:30458165"
/sex="male"
/dev_stage="embryonic day 8"
/lab_host="DH10B"
/clone_lib="NIA Mouse Embryonic Germ Cell cDNA library (long)"
/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Dr. Mark G. Carter (NIH/NIA-IRP). EG cells were cultured at 37. C, 5% CO2 in DMEM supplemented with 15% ES cell-qualified FBS, 0.1mM non-essential amino acids, 2 mM glutamine, penicillin/streptomycin, 1 mM sodium pyruvate, 0.1 mM beta-mercaptoethanol, and 10000000 units of LIF per liter. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTCTTGTAGTCGAGCGCGCCCTTTT-3'] from 2.5 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 4.0 kb. The library was constructed by Yulan Piao."

ORIGIN

Query Match 4.1%; Score 63.2; DB 6; Length 485;
Best Local Similarity 67.4%; Pred. No. 0.0016;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 533 GAGCAGTGTGCTGTGTCGAGCGGAGCTATCTGTGACGGCTCGCCTCG 592
DB 55 GGAATCTGTGCGCTGCAAGAGCGGAGGCTCTGCTGTGACACATGCCCTTC 114
QY 593 GGCCTTCACCTGAGCTGCTGTCCTCGCTCGGAGATCCCACTGGAG 652
DB 115 TTCTTACACATCCTGAGTGAACCCCGCTGCCAGATCCCAACGGGCAATGGCT 174

QY 653 GGTCTCCAGCTG 664
DB 175 GTGTCCCGCTG 186

RESULT 29
CA893864
LOCUS
DEFINITION

CA893864 514 bp mRNA linear EST 20-DEC-2002
B0181H02-5N NIA Mouse Neural Stem Cell (Differentiated) cDNA
Library (long) Mus musculus cDNA clone NIA:B0181H02 IMAGE:30101653
5', mRNA sequence.

ACCESSION
CA893864

CA893864.1 GI:27345413

KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 514)
Piao, Y., Dudekula, D. B., Qian, Y., Martin, P. R., Aiba, K., Vescevi, A. L. and Ko, M. S. H.
Systematic Analyses of NIA Mouse Neural Stem Cell (Differentiated) cDNA Library (long)
Unpublished (2002)

JOURNAL

COMMENT

Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Caswell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: B0181 row: H column: 02
Seq primer: -21M13 Reverse
High quality sequence stop: 514
POLYA-No.

FEATURES
source

1. 514
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CD1"
/db_xref="niaBST:B0181H02-5N"
/db_xref="taxon:10090"
/clone="NIA:B0181H02 IMAGE:30101653"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="NIA Mouse Neural Stem Cell (Differentiated) cDNA Library (long)"
/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Dr. Angelo L. Vescevi (Institute for Stem Cell Research, Italy). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTCTTGTAGTCGAGCGCGCCCTTTT-3'] from 2.0 microgram of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.2 kb. The library was constructed by Yulan Piao."

ORIGIN

Query Match 4.1%; Score 63.2; DB 6; Length 514;
Best Local Similarity 67.4%; Pred. No. 0.0016;

	Matches	89;	Conservative	0;	Mismatches	43;	Indels	0;	Gaps	0;
OY	513	GGAGCAGTGTGCGCTGTGCGGAGACGGGGGAGCTCATTCGTCTGTAGACGGCTGCCCTCG	592							
Db	312	GGAAATTCGTGTGCGCTGTGCGGAGACGGGGGAGCTCCTGTGCTGTGACACATGCGCCCTTC	371							
OY	593	GGCCTTCCACGTCGCGCTGTGCCCTCCGCTCCGGAGATGCCCAATGGAGCCTGGAG	652							
Db	372	TTTCTACACATTCACCTGCTGAAACCCCCCGCTGCAGAGATCCCAACGGGGAATGGCT	431							
OY	653	GTGCTCCAGCTG	664							
Db	432	CTGTCCCGCTG	443							

RESULT 30					
CAS64994					
LOCUS	CAS64994	562 bp	mRNA	linear	EST 19-NOV-2002
DEFINITION	K0332G02-5N NIA Mouse Osteoblast cDNA library (Long) Mus musculus				
	cDNA clone NIA:K0332G02 IMAGE:30057385 5', mRNA sequence.				
ACCESSION	CAS64994				

VERSION	CAS64994.1	GI:25109673
KEYWORDS	EST.	
SOURCE	Mus musculus	(house mouse)
ORGANISM	Mus musculus	

REFERENCE
1 (bases 1 to 562)
AUTHORS
Piao, Y., Kargul, G. J., Dudekula, D. B., Qian, Y., Luo, A., Carter, M. G.

TITLE	Systematic Analyses of NIA Mouse Osteoblast cDNA Library (Long
JOURNAL	Unpublished (2001)
COMMENT	Other_ESTB: K0332G02-3

Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.gsc.nia.nih.gov
Plate: K0332 Row: G Column: 02
Seq primer: M13 Reverse
High quality sequence stop: 562
POLYA=NO.

FEATURES	Location/Qualifiers
Source	1. .562

```

/organism="Mus musculus"
/mol_type="mRNA"
/strains="C3H/He"
/db_xref="nleST:K0332602-5N"
/db_xref="taxon:10090"
/cclone="NTA:K0332602 IMAGE:30057385"
/tissue_type="Oesophagus"
/cell_line="KUSA-A1 cells"
/lab_host="DH10B"

```

11:1353-1558 (2001). [PMID: 11544191]. Total RNAs were obtained from Dr. Akihito Umezawa (Keio University School of Medicine, Japan). Double-stranded cDNAs were synthesized with an Oligo(dT) primer (Invitrogen: 5'-pGACTAGTCTTCAGATCCGACGGCCGCTTTTCTTTT-3') from 2.1 µg of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker *lu*-SalI, purified by phenol/chloroform, and separated from free linkers by Centrificon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerases (Takara) with a primer SalI-5'. The products were purified by phenol/chloroform and Centrificon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector.

ORIGIN

Query Match	4.1%;	Score 63.2;	DB 6;	Length 562;
Best Local Similarity	67.4%;	Pred. No. 0.0016;		
Matches 89;	Conservative 0;	Mismatches 43;	Indels 0;	Gaps 0;

Qy	533	GGACGAGTGTGCGGTGTGTGCGGAGCGGGGGGAGCTATCTGTGTGACGGCTCCCTCG	592
Db	73	GGAAATCTCTCCGGTGTGCAGAGAGGGGGGAGCTCTGTGTGTGACATATCCCTTC	132
Qy	593	GGGCTTCAACTGAGCTGTGCTGTCCCTTCGCTCCGGAGATCCCACTGGGACCTGAG	652
Db	133	TTCCTACCAATCACTGCTGTGAAACCCCGCTGCAGATGCCCAACGGCGAATGCT	192
Qy	653	GTGCTCCAGCTG	664
Db	193	CTGTCCCGCTG	204

RESULT 31
BB654405

LOCUS BB654405 564 bp mRNA linear EST 26-OCT-2001
DEFINITION BB654405 RIKEN full-length enriched, 2 days neonate thymus thymic
cells Mus musculus cDNA clone C920008E20 5', mRNA sequence.

ACCUSATION	BB654405.1	GI:16498233
VERSION		
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	

REFERENCE AUTHORS

TITLE	JOURNAL	COMMENT
Unpublished (2001)		
Contact: Yoshihide Hayaehizaki		
RIKEN Mouse ESTs (Aitakawa, T., et al. 2001)		
		<p> Hara, A., Hiramoto, K., Horii, F., Ihnly, Y., Ito, M., Kawai, J., Komno, H., Kouda, M., Koya, S., Matsumura, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shingura, A., Shiraki, T., Sugabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takekida, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayaehizaki, Y. </p>

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)
wagi, K., Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubuchi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
Kondo, S., Shitagawa, A., Saito, T., Kiyosawa, H., Yamamata, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

/clone="NIA:B0325H01 IMAGE:30431508"

5'-pGACGTCTTCATGATGCGACGCGCGCCGCTTTTCTTTTCTTTT-3'] from 6.1 µg of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L⁺-SalI, purified by phenol/chloroform, and separated from free linkers by Centricion 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricion 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3 kb. The library was constructed by Yulan Piao."

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: PYX-5.

FEATURES

source

1..625

Location/Qualifiers

/organism="Mus musculus"

/mol_type="rRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:30630778"

/issue_type="whole eye"

/dev_stage="embryo 12.5,13.5,14.5 dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_1ib="NIH BMAP_HB0"

/note="Organ: Eye; Vector: PYX-Asc; Site:1: EcoR I; Site:2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTTATGAGCT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Query Match

Best Local Similarity 4.1%; Score 63.2; DB 7; Length 625;

Best Local Similarity

67.4%; Pred. No. 0.0016; Mismatches 43; Indels 0; Gaps 0;

Matches

89; Conservative 0;

Query

533 GGACGAGTGTGCGTGTGCGGAGCGGAGCTCATCTGTGACGCGCTCG 592

Database

179 GGAATCTCTCGGTGTGCAAGAGCGCGGAGCTCTGTGTGACACATGCCCTTC 238

Query

593 GGCCTTCCACCTGCGCTGCTCCCTCCCTCCGCGGAGATCCCAATGGAGCT 652

Database

239 TTCTTACCACTCATCTGCTGCAACCCCGCTGCGAGATCCCAAGCGGAGTGGCT 298

Query

653 GTGCTCCAGCTG 664

Database

299 CTGTCCCGCTG 310

Db 299 CTGTCCCGCTG 310

RESULT 34

BU705998

LOCUS

634 bp mRNA linear EST 15-JUL-2003

DEFINITION

UT-M-FOO-caa-1-10-0-UI.r1 NIH_BMAP_F00 Mus musculus cDNA clone

IMAGE:6408657 5', mRNA sequence.

ACCESSION

BU705998

VERSION

BU705998.1

KEYWORDS

GI:23636034

SOURCE

EST.

ORGANISM

Mus musculus (house mouse)

REFERENCE

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 634)

TITLE

NIH-MGC http://mgi.nci.nih.gov/

JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT

Unpublished (1999)

CONTACT

Robert Strauberg, Ph.D.

Email: cgaab-r@mail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/BLN at:

http://image.llnl.gov

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES

source

1..634

Seq primer: PYX-5.

Location/Qualifiers

/organism="Mus musculus"

/mol_type="rRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:6408657"

/issue_type="whole brain"

/dev_stage="embryo 12.5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_1ib="NIH BMAP_F00"

/note="Organ: Brain; Vector: PYX-Asc; Site:1: EcoR I; Site:2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAAGAGCC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN

Query Match

Best Local Similarity 4.1%; Score 63.2; DB 5; Length 634;

Best Local Similarity

67.4%; Pred. No. 0.0016; Mismatches 43; Indels 0; Gaps 0;

Matches

89; Conservative 0;

Query

533 GGACGAGTGTGCGTGTGCGGAGCGGAGCTCATCTGTGACGCGCTCG 592

Database

269 GGAATCTCTCGGTGTGCAAGAGCGCGGAGCTCTGTGTGACACATGCCCTTC 328

Query

593 GGCCTTCCACCTGCGCTGCTCCCTCCCTCCGCGGAGATCCCAATGGAGCT 652

Database

329 TTCTTACCACTCATCTGCTGCAACCCCGCTGCGAGATCCCAAGCGGAGTGGCT 388

Query

653 GTGCTCCAGCTG 664

Database

389 CTGTCCCGCTG 400

Db 389 CTGTCCCGCTG 400

RESULT 35

BU055310

LOCUS

635 bp mRNA linear EST 26-AUG-2002

DEFINITION

UT-M-FOO-bap-b-16-0-UI.r1 NIH_BMAP_F00 Mus musculus cDNA clone

IMAGE:6405423 5', mRNA sequence.

ACCESSION

BU055310

VERSION

BU055310.1

KEYWORDS

GI:22495387

SOURCE

EST.

ORGANISM

Mus musculus (house mouse)

REFERENCE

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 635)

TITLE

NIH-MGC http://mgi.nci.nih.gov/

JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT

Unpublished (1999)

CONTACT

Robert Strauberg, Ph.D.

Email: cgaab-r@mail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/BLN at:

http://image.llnl.gov

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES Seq primer: pyx-5.
Location/Qualifiers
source 1..635
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6405423"
/issue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="NIH BMAP F00"
/note="Organ: Brain; Vector: pyx-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TAGAGAGCC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN
Query Match 4.1%; Score 63.2; DB 5; Length 635;
Best Local Similarity 67.4%; Pred. No. 0.0016;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 533 GGACGAGTGTGCGCTGTGTCGGAGCGCGGAGCTCATCTGCTGACGCGCTCG 592
DB 14 GGATTCTGTCCGCTGTGACAGCGCGGAGCTCTGTGTCGACACATGCCCTTC 73
QY 593 GGCTTCACACCTGAGCTGTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGAGCTGAG 652
DB 74 TTCTTACCACTCATCTGCTGAACCCCGCTGCCAGAGATCCCAAGCGCAATGGCT 133
QY 653 GTGCTCCAGCTG 664
DB 134 CTGTCCCGCTG 145

RESULT 36
CB246896 723 bp mRNA linear EST 09-JUL-2003
LOCUS UI-M-FIO-cdy-k-21-0-UI.r1 NIH BMAP_FIO Mus musculus cDNA clone
DEFINITION IMAGE:6836206 5', mRNA sequence.
ACCESSION CB246896
VERSION CB246896.1 GI:28368540
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 723)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES Seq primer: pyx-5.
Location/Qualifiers
source 1..723
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6836206"
/issue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="NIH BMAP F10"
/note="Organ: Brain; Vector: pyx-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCAGAC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN
Query Match 4.1%; Score 63.2; DB 6; Length 723;
Best Local Similarity 67.4%; Pred. No. 0.0017;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 533 GGACGAGTGTGCGCTGTGTCGGAGCGCGGAGCTCATCTGCTGACGCGCTCG 592
DB 357 GGATTCTGTCCGCTGTGACAGCGCGGAGCTCTGTGTCGACACATGCCCTTC 416
QY 593 GGCTTCACACCTGAGCTGTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGAGCTGAG 652
DB 417 TTCTTACCACTCATCTGCTGAACCCCGCTGCCAGAGATCCCAAGCGCAATGGCT 476
QY 653 GTGCTCCAGCTG 664
DB 477 CTGTCCCGCTG 488

RESULT 37
CF726532 730 bp mRNA linear EST 09-OCT-2003
LOCUS UI-M-HB0-cdh-h-09-0-UI.r1 NIH BMAP_HB0 Mus musculus cDNA clone
DEFINITION IMAGE:30547664 5', mRNA sequence.
ACCESSION CF726532
VERSION CF726532.1 GI:37600700
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 730)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/mousefl.html>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pyx-5.

FEATURES

source

Location/Qualifiers

1..730

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:30547664"

/issue_type="whole eye"

/dev_stage="embryo 12.5,13.5,14.5 dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_1ib="NIH BMAP H30"

/note="Organ: Eye; Vector: pYX-Asc; Site:1: EcoR I; Site:2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTAATGAGT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Query Match

Best Local Similarity 4.1%; Score 63.2; DB 7; Length 730; Pred. No. 0.0017;

Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGAGTGTCCGCTGTGTCGGAGCGCGGAGCTCATCTGTGTGACGCGCTCG 592
DB 51 GGAAATTCCTCGGCTGTGACAGAGCGCGGAGCTCTGTGTGTGACACATGCCCTTC 110
QY 593 GGCGTTCCACCTGCGCTGCTGCCCTCGCTCGCGGAGATCCCGAGTGGAGCTGGAG 652
DB 111 TTCTTACCACTCATCTGCTGTAACCCCGCTGCCAGAGATCCCAAGCGCAATGGCT 170
QY 653 GTGCTCCAGCTG 664
DB 171 CTGTCCCGCGCTG 182

RESULT 38

CNS34827

LOCUS

DEFINITION UI-M-HSO-cgd-k-04-0-UI.r1 NIH_BMAP_HSO Mus musculus cDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

736 bp mRNA linear EST 29-APR-2004

IMAGE:30674835 5', mRNA sequence.

CNS34827

CNS34827

EST.

Mus musculus

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 736)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabbs-remail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/mouseefl.html

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seg primer: pYX-5.

Location/Qualifiers

1..736

FEATURES

source

ORIGIN

Query Match

Best Local Similarity 4.1%; Score 63.2; DB 7; Length 736; Pred. No. 0.0017;

Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GGACGAGTGTCCGCTGTGTCGGAGCGCGGAGCTCATCTGTGTGACGCGCTCG 592
DB 46 GGAAATTCCTCGGCTGTGACAGAGCGCGGAGCTCTGTGTGTGACACATGCCCTTC 105
QY 593 GGCGTTCCACCTGCGCTGCTGCCCTCGCTCGCGGAGATCCCGAGTGGAGCTGGAG 652
DB 106 TTCTTACCACTCATCTGCTGTAACCCCGCTGCCAGAGATCCCAAGCGCAATGGCT 165
QY 653 GTGCTCCAGCTG 664
DB 166 CTGTCCCGCGCTG 177

RESULT 39

CO045219

LOCUS

DEFINITION UI-M-HO0-cmx-0-19-0-UI.r1 NIH_BMAP_HO0 Mus musculus cDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

754 bp mRNA linear EST 10-JUN-2004

IMAGE:30638082 5', mRNA sequence.

CO045219

CO045219

EST.

Mus musculus

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 754)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabbs-remail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/mouseefl.html

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seg primer: pYX-5.

Location/Qualifiers

1..754

/organism="Mus musculus"

/mol_type="mRNA"

FEATURES

source

```

/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30638082"
/tissue_type="Upper Head"
/dev_stage="9.5-10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP H00"
/notes="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

```

ORIGIN

```

Query Match      4.1%; Score 63.2; DB 7; Length 754;
Best Local Similarity 67.4%; Pred. No. 0.0017;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```

```

QY 533 GGACGAGTGTGCGGTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCG 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 39 GGAATTTCTTCCGCTGTGCAAGAGACGCGGAGCTCTGTGTGACACATGCCCTTC 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 593 GGCCTTCCACCTGCGCTGTGCGCTCGCTCGGAGATCCCGATGGAGACTGAG 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99 TTCTTACCAATCCATCTGCTGACACCCCGCTGCGAGATCCAAAGCGGAATGGCT 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 653 GTGCTCCAGCTG 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 159 CTGTCCCGCGCTG 170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 40
LOCUS CN536888 755 bp mRNA linear EST 29-APR-2004
DEFINITION UI-M-HOO-cqb-j-11-0-UI.r1 NIH BMAP_HOO Mus musculus cDNA clone
IMAGE:30664450 5', mRNA sequence.
ACCESSION CN536888
VERSION CN536888.1 GI:46865044
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 755)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Email: cgaabrs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

```

```

FEATURES
source 1..755
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"

```

```

/clone="IMAGE:30664450"
/tissue_type="Upper Head"
/dev_stage="9.5-10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP H00"
/notes="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

```

ORIGIN

```

Query Match      4.1%; Score 63.2; DB 7; Length 755;
Best Local Similarity 67.4%; Pred. No. 0.0017;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```

```

QY 533 GGACGAGTGTGCGGTGTGCGGAGCGGCGGAGCTCATCTGTGACGCGCTCG 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 46 GGAATTTCTTCCGCTGTGCAAGAGACGCGGAGCTCTGTGTGACACATGCCCTTC 105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 593 GGCCTTCCACCTGCGCTGTGCGCTCGCTCGGAGATCCCGATGGAGACTGAG 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 106 TTCTTACCAATCCATCTGCTGACACCCCGCTGCGAGATCCAAAGCGGAATGGCT 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 653 GTGCTCCAGCTG 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 166 CTGTCCCGCGCTG 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 41
LOCUS CA327845 762 bp mRNA linear EST 09-UTL-2003
DEFINITION UI-M-FYO-ccw-f-03-0-UI.r1 NIH BMAP_FYO Mus musculus cDNA clone
IMAGE:6825316 5', mRNA sequence.
ACCESSION CA327845
VERSION CA327845.1 GI:24545943
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 762)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Email: cgaabrs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIN at:
http://image.jlini.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

```

```

FEATURES
source 1..762
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6825316"

```

/cissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP FY0"
/note="Organ: Brain; Vector: pYX-Aac; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction. ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Aac vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
program coordinator."

ORIGIN

Query Match 4.1%; Score 63.2; DB 6; Length 762;
Best Local Similarity 67.4%; Pred. No. 0.0017;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GAGAGAGTGTCCGCTGTGCGGAGCGCGGAGCTCATCTGTGACGGCTGCCCTCG 552
DB 342 GGAATTCTGTGCGCTGTGCAAGGAGCGGCGGAGCTCTGTGCTGTGACACATGCCCTTC 401
QY 593 GGGCTTCACCTGCGCTGTGCTGCCCTCGCTCCGCGGAGATCCCGACGTGGACCTGGAG 652
DB 402 TTCTTACACATCTCAGCTGTGAACCCCGCTGTCCAGAGATCCAAACGGCGAATGGCT 461
QY 653 GTGCTCCAGCTG 664
DB 462 CTGTCCCGCTG 473

RESULT 42
CA749488 768 bp mRNA linear EST 09-JUL-2003
LOCUS UI-M-FYO-cdd-n-01-0-UI.r1 NIH_BMAP_FYO Mus musculus cDNA clone
DEFINITION IMAGE:6831650 5', mRNA sequence.
ACCESSION CA749488
VERSION CA749488.1 GI:25571583
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

source location/Qualifiers
1..768
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6831650"

/cissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP FY0"
/note="Organ: Brain; Vector: pYX-Aac; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction. ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Aac vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
program coordinator."

ORIGIN

Query Match 4.1%; Score 63.2; DB 6; Length 768;
Best Local Similarity 67.4%; Pred. No. 0.0017;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 533 GAGAGAGTGTCCGCTGTGCGGAGCGCGGAGCTCATCTGTGACGGCTGCCCTCG 552
DB 97 GGAATTCTGTGCGCTGTGCAAGGAGCGGCGGAGCTCTGTGCTGTGACACATGCCCTTC 156
QY 593 GGGCTTCACCTGCGCTGTGCTGCCCTCGCTCCGCGGAGATCCCGACGTGGACCTGGAG 652
DB 157 TTCTTACACATCTCAGCTGTGAACCCCGCTGTCCAGAGATCCAAACGGCGAATGGCT 216
QY 653 GTGCTCCAGCTG 664
DB 217 CTGTCCCGCTG 228

RESULT 43
CK637900 787 bp mRNA linear EST 28-JAN-2004
LOCUS UI-M-HO0-cnt-m-12-0-UI.r1 NIH_BMAP_HO0 Mus musculus cDNA clone
DEFINITION IMAGE:30642251 5', mRNA sequence.
ACCESSION CK637900
VERSION CK637900.1 GI:41363766
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mouse1.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

source location/Qualifiers
1..787
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30642251"
/cissue_type="Upper Head"

Db 529 MYSSSVSCCSCGCTCCGCGCTGCTGCKGCGCGCTSCSSSSCCSBBSYST 588
 QY 935 TGGTGGCGCTTGGGGGTGGGAGATGGTAGCGATGCTGGGTACTGACACGCG 994
 Db 589 CCGTCTCTKCSGCGTSGCTGCTGCGGAGCGCGGCGCGGCGCGGCGCGGCGGCGG 648
 QY 995 CGCTGCTTCCACATGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 1054
 Db 649 SSGSSSSSSCGSS 708
 QY 1055 GCGCTGCAATCTCTGCTGAGAGACGTGACCTCCAGCCCTGCTGAGAGGGTCTGCGCCC 1114
 Db 709 SGGGCGSCGCGSCGCGSCGCGSCGCGSCGCGSCGCGSCGCGSCGCGSCGCGSCG 768
 QY 1115 CAGCCCG 1174
 Db 769 CGGSSCGGCG 828
 QY 1175 TCTGCAACAGGATGATGCTGAGATCTCTTCTGAGCGAGACACCTTCGATGCA 1234
 Db 829 CGGCG 888
 QY 1235 GTGGCGCATTCAGAGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1280
 Db 889 GCGSGGCGGCG 934

RESULT 48
 CR715927
 LOCUS CR715927 1066 bp mRNA linear HTC 12-AUG-2004
 DEFINITION Tetradon nigroviridis full-length cDNA.
 CR715927
 VERSION CR715927.1 GI:51214161
 KEYWORDS HTC; cDNA; full-length; Tetradon nigroviridis.
 SOURCE Tetradon
 ORGANISM Tetradon
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodontidae.
 1 (bases 1 to 1066)
 Genoscope.
 AUTHORS Direct Submission
 JOURNAL Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT The sequences are based on single pass reads.
 More information available at
 http://www.genoscope.cns.fr/tetradon.
 FEATURES
 source
 1. 1066
 /organism="Tetradon"
 /mol_type="mRNA"
 /db_xref="taxon:47144"
 /tissue_type="Eggs"

ORIGIN
 Query Match 4.1%; Score 63.2; DB 3; Length 1066;
 Best Local Similarity 56.9%; Pred. No. 0.0017;
 Matches 116; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 546 GTGTGTGGGAGCGCGGAGCTCATCTGCTGTAAGCGGCTCGGCGCTTTCACCTG 605
 Db 229 GTGTGCAAGATGAGAGAGAGCTTGTGCTGTGACACCTGCGCTCTCTACACATC 288
 QY 606 GCGTGTCTGCTCTCGCTCGCGGAGATCCCAAGTGGAGCTGCTCCAGCTGC 665
 Db 289 CACTGCTCAACCTCTCCCTCCCGAATCCCTATGAGAAATGAGATCTGCCCCGCTGC 348
 QY 666 CTGACGCAACAGTCCGAGAGGTGCACCCCGGCAAGAGAGCCCGCCCGAGAGCA 725
 Db 349 AAGGTTCACCAATGAGAGGGAAGTTTCAAGAAATTTTAACTTGGCGATGGGGGACCA 408

QY 726 CCCGTGAGACCCCGCTCCCCCG 749
 Db 409 CCCGCCCCACACTGTCTCTCGG 432

RESULT 49
 AK034549
 LOCUS AK034549 3137 bp mRNA linear HTC 03-APR-2004
 DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region
 and neck cDNA. RIKEN full-length enriched library, clone:943004K15
 product:CHROMODOMAIN HELICASE-DNA-BINDING PROTEIN 4 (CHD-4) (MI-2
 AUTOMANTIGEN 218 KDA PROTEIN) (MI2-BETA) homolog [Homo sapiens],
 full insert sequence.

ACCESSION AK034549
 VERSION AK034549.1 GI:26330020
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Komano, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 11042159
 PUBMED 10493374

REFERENCE
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Komano, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system -384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE
 AUTHORS 6 (bases 1 to 3137)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kigawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komano, H., Konda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohnato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission
 JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

COMMENT

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>
Location/Qualifiers

FEATURES

source

```
1.3137
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:9430004K15"
/db_xref="taxon:10090"
/clone="9430004K15"
/tissue_type="embryonic body between diaphragm region and neck"
/clone_1ib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
<1.3137
/note="unnamed protein product; CHROMODOMAIN HELICASE-DNA-BINDING PROTEIN 4 (CHD-4) (MI-2 AUTOANTIGEN 218 KDA PROTEIN) (MI2-BETA) homolog [Homo sapiens] (SWISSPROT|Q14839, evidence: PASTY, 99.8%ID, 54.6%length, match=3135) putative"
/codon_start=3
/protein_id="BAC28749.1"
/db_xref="GI:26330021"
/translation="PIRKAKTEGKGPNNRRKSGSPVPADAKPKKVAIPKIKLG GSGKRKSSSEDDLDVSDPDDASINSTSVDSSTSRSRKRLRDKKKKKEE EYAVADVGETDHDQCEVCOQGEIILCDPCPAYMVCIDPMEKAPRQKSCPEE KEGIOWEAKEDNESEGEIIELEVGDPEEDDHMEFCVCKDGEILCCDTPSSYHI HCLNPLPEIPIENEMLCPRCTCPALGKVKOKLIMKGGQSPPTPVPRPDAPNPS PKLEBRPEKOPFVKWQGSYHCVSWSELOELHCOVMPRNVORNDMDPEPSGPG GDBEKSRKKNQPKPAEMBERPYRGIRPEMMITRILNHSYDKGHYHILKRDLP IDQASWESDVEIIDIYDLFKOSYMHRELHGESEKPGKLLKVKVLRKLERPEIPT VPTVYERQPEIADTGTGLHRYQWGLMWLFSAQGDITLADMGKLEKTPQAV PYSLYKEGSKGPFVLSAPLSTIIMBERFEKMAADMYVTIVGDKDSRAIIRENEP SFEDNAIRGGKSKRMKEASVKEFVILTSYELITDMAIIGSIDMACILVDEARLK NNSKREFVLNGYSLOHKLILGTPLONNLELPHLPLTPEPHNLGSEEPADI AKEDQIKKLDHNLGPHMLRLKADVFRKNPSKTELVRYELSMOKYKTYITITRPE ALVARGGQNOVSLANVMDKCCNHPYLPVAMEAPKPNMGYDGSALITRASGLL LQKMLKNLKEGGHRLVIPSQMTKMLDLEDFLEHGYKTERIDGQITGMRBEALDR FNAFGAOQCFGLSTRAIGIIMATADTVIIVDSQWPNNDIQASRAHRIQONKRV MYRFTVTRASVEERTIOVAKKQMLTHLVVRLPGISKTSMSQEDLDILKFTTELF KDEATVGGDKNGEDSVLHYDKAIERLLDNODETBELQGNNEVLSKFAVOY VREEMGESEEVER"
```

CDS

```
ORIGIN
Query Match 4.1%; Score 63.2; DB 3; Length 3137;
Best Local Similarity 67.4%; Pred. No. 0.002;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
```

```
QY 533 GAGCAGAGTGCCTGTGTGGGACGGCGGGAGACTCATCTGCTGACGGCTGCCG 592
DB 584 GGAATTCCTGTGGGTCTGCAAGAGCGCGGGAGCTCTCTGTGTCACATGACCTTC 643
QY 593 GGCCTTCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATCCCGAGTGGAGCTGGAG 652
DB 644 TTCTTACCATCATCTGCTGAACCCCGCTGCCAGATCCCAAGCGCAATGGCT 703
QY 653 GTGCTCAAGCTG 664
DB 704 CTGTCCCGCTG 715
```

RESULT 50

LOCUS

CNS0091P/c 925 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION

AL053013 GI:4934461

VERSION

GSS.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster (fruit fly)

REFERENCE

Neupert, Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

AUTHORS

Genoscope.

TITLE

Direct Submission

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw bp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

```
1.925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_1ib="RPCI-98"
/note="end : TET3"
```

ORIGIN

```
Query Match 4.1%; Score 62.8; DB 9; Length 925;
Best Local Similarity 13.3%; Pred. No. 0.0021;
Matches 51; Conservative 181; Mismatches 152; Indels 0; Gaps 0;
```

```
QY 436 GCTGGGGAGCTGTTTGGAGAGAGGAGCTCTCAGAGGAGTGCACCCAGCCAG 495
DB 916 SCSSBSCSSSMSTSSNSBSCSSSSSSSTSSMSSSSSSSSSGTSACV 857
QY 496 TCTGCATGCGCCTCTTCTCTGTCAGAGAAGATGAGACAGAGTGCCTGTGCGG 555
DB 856 KGNASSSCCGCGGABCCMCSSSSCCGSASRGVNVASGAGRGGSGSAGSASHS 797
QY 556 ACGGCGGAGAGCTCATCTGCTGACGCTGCGCTTCCTTCCACTGAGCTGCTGCT 615
DB 796 SAGCBSSSSSCASCMASSSSSSASRSRSGGAGGASASRSASAGSVAS 737
QY 616 CCCCTCGCTCCGGAGATCCCGAGTGGAGCTGAGAGTGTCTCAGCTGCTGACGCA 675
DB 736 ASSSSSSSSSSVSSSVASMSCSBSSSSASASSSSSSSSASCSGCTTWSGSCST 677
QY 676 CAGTCAGAGAGTGCAGCCCGGAGAGAGAGCCCGGCGGAGCCAGCCAGCCAGAGA 735
DB 676 SASMSASRSSSSSSSSSSSSSASASSSSSSSSSSSSSSSSAGSAGSGG 617
QY 736 CCCGCTCCCGCGGGCTTAAAGTCCGCGGAGAGAGAGTAAAGTCACTGAGGAG 795
DB 616 SVASASGMSVSSSGGSGGSGGCGGCGGCGGSGSSSSSSSGGSGGSGGSGGSGG 557
```

QY 796 CCTACCGCATGACAGCATC 819
DB 556 SSAAAAAASCVASCGMCGSKS 533

RESULT 51
LOCUS BGS11885
DEFINITION BGS11885 587 bp mRNA linear EST 28-MAR-2001
dad29d08.y1 wellcome CRC PCS107 tropicalis St10-12 Xenopus
tropicalis cDNA clone IMAGE:4440999 5' similar to TR:095854 095854
TRANSCRIPTIIONAL INTERMEDIARY FACTOR 1 ALPHA. ;, mRNA sequence.

ACCESSION BGS11885

VERSION BGS11885.1 GI:13482542

KEYWORDS EST

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 587)
Clifton, S., Johnson, S. L., Blumberg, B., Song, J., Hillier, L.,
Pape, D., Martin, J., Wyle, T., Underwood, K., Theising, B., Bowers, Y.,
Peterson, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterson, R. and Wilson, R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Library constructed by A. Zorn and J. Mason (Wellcome/CRC
Institute). DNA Sequencing by: Washington University Genome
Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LNL at: infoimage.lnl.gov
Seq primer: -40RP from Glibco
High quality sequence stop: 502.

FEATURES

source

1. 587
Location/Qualifiers
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8354"
/clone="IMAGE:4440999"
/issue_type="whole embryo, stages 10-12"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Wellcome CRC PCS107 tropicalis St10-12"
/note="Vector: PCS107; Site 1: NotI; Site 2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Average
insert size 1.5 kb, range 0.5-4 kb. Library constructed by
A. Zorn and J. Mason (Wellcome/CRC Institute)."

ORIGIN

Query Match 4.1%; Score 62.6; DB 4; Length 587;

Best Local Similarity 61.2%; Pred. No. 0.0022;

Matches 101; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 528 AATGAGACGAGTGTCCGTGTGCGGAGCGCGGAGCTCATCTGTGTGACGCGCTCG 587
DB 423 AATGAGACGAGTGTGCTGTATGTATTAATGCGGGGAGCTGCTTTGCTGTGAAATGT 482
QY 588 CCTCGGGCTTCCACTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGGACC 647
DB 483 CCCAAAGTTTCCACCTGCTGTGTGCGACGTCCTACCTGATGAACCTTCCCACTGTGTGA 542
QY 648 TGGAGGNGCTCCAGCTGCTGCGGCAAGTCACAGAGGTGAG 692
DB 543 TGGATCTGACCTTTTGTCCGGGACCTGTGACAGCCAGAGTGGAG 587

RESULT 52

LOCUS CN526117

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CNS26117 677 bp mRNA linear EST 29-APR-2004
UI-M-HNO-coc-j-05-0-UI-r1 NIH_BMAP_HNO Mus musculus cDNA clone
IMAGE: 30654076 5', mRNA sequence.
CN526117
CN526117.1 GI:46853770
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 677)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mouse1.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.

FEATURES

source

Location/Qualifiers
1. 677
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 30654076"
/issue_type="Upper Head"
/dev_stage="9.5-10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP HNO"
/note="Organ: Head; Vector: pYX-Anc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Anc vector. The library tag
sequence located between the Not I site and the polyA tail
is CCAAGTGAAT. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

ORIGIN

Query Match 4.0%; Score 62.2; DB 7; Length 677;

Best Local Similarity 66.7%; Pred. No. 0.0028;

Matches 88; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 533 GAGACAGTGTCCGTGTGCGGAGCGCGGAGCTCATCTGTGTGACGCGCTCG 592
DB 506 GGAATTTCTTCTCCGCTGTGAGAGCGCGGAGCTCCTGTGTGACACATGCTTC 565
QY 593 GGCCTTCCACCTGGGCTGCTGCCCTCCGCTCCGGAGATCCCAAGTGGAGCTGAG 652
DB 566 TTTCATACCATTCATCAGCTGCTGAACCCCGCTGCGACAGATCCANAGCGAATGACT 625
QY 653 GTGCTCCAGCTG 664
DB 626 CTGTCCCGCTG 637

RESULT 53

CBS23529

LOCUS CB523529 680 bp mRNA linear EST 09-JUL-2003
 DEFINITION UI-M-GHO-cer-1-07-0-UI.r1 NIH_BMAP_GHO Mus musculus cDNA clone
 IMAGE:643512 5', mRNA sequence.
 ACCESSION CB523529
 VERSION CB523529.1 GI:29356884
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 680)
 NIH-MGC <http://imgc.ncl.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 TITLE
 AUTHORS
 REFERENCE
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 DNA Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/mouseefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
 Seq primer: PYX-5.
 Location/Qualifiers
 source
 1..680
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:643512"
 /issue_type="Whole brain"
 /dev_stage="1, 5, and 15 days newborn"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_1ib="NIH BMAP GH0"
 /note="Organ: Brain; Vector: PYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTCAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 4.0%; Score 62.2; DB 6; Length 680;
 Best Local Similarity 66.7%; Pred. No. 0.0028; Indels 0; Gaps 0;
 Matches 88; Conservative 0; Mismatches 44;
 QY 533 GGACGAGTGTGCGGTGTGCGAGCGGAGGAGCTCATCTGCTGTGACGCGCTCGCCCTG 592
 DB 482 GGAATCTGTCCGCTGTGCAAGAGCGGCGGAGCTCTGTGTCGACACATGCCCTTC 541
 QY 593 GGCCTTCACCTGCGCTGTGCTCCCTCGCTCCGCGAGATCCCAAGTGGAGCTGGAG 652
 DB 542 TTCCTACACATTCACCTGCTGAACCCCGCTGCCAGAGATCCAAACGGCAATGGCT 601
 QY 653 GTGCTCCAGCTG 664
 DB 602 CTGTCCCGCTG 613

RESULT 54
 BU056500/c
 LOCUS BU056500

751 bp mRNA linear EST 26-AUG-2002

DEFINITION UI-M-F00-cab-h-16-0-UI.r1 NIH_BMAP_F00 Mus musculus cDNA clone
 IMAGE:6409023 5', mRNA sequence.
 ACCESSION BU056500
 VERSION BU056500.1 GI:22496577
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 751)
 NIH-MGC <http://imgc.ncl.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 TITLE
 AUTHORS
 REFERENCE
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 DNA Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
 Seq primer: PYX-5.
 Location/Qualifiers
 source
 1..751
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6409023"
 /issue_type="Whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_1ib="NIH BMAP F00"
 /note="Organ: Brain; Vector: PYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAGAGGCC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 4.0%; Score 62.2; DB 5; Length 751;
 Best Local Similarity 66.7%; Pred. No. 0.0028; Indels 0; Gaps 0;
 Matches 88; Conservative 0; Mismatches 44;
 QY 533 GGACGAGTGTGCGGTGTGCGAGCGGAGGAGCTCATCTGCTGTGACGCGCTCGCCCTG 592
 DB 630 GGAATCTGTCCGCTGTGCAAGAGCGGCGGAGCTCTGTGTCGACACATGCCCTTC 571
 QY 593 GGCCTTCACCTGCGCTGTGCTCCCTCGCTCCGCGAGATCCCAAGTGGAGCTGGAG 652
 DB 570 TTCCTACACATTCACCTGCTGAACCCCGCTGCCAGAGATCCAAACGGCAATGGCT 511
 QY 653 GTGCTCCAGCTG 664
 DB 510 CTGTCCCGCTG 499

RESULT 55
 AG561783
 LOCUS AG561783

776 bp DNA linear GSS 05-JUN-2004

DEFINITION Mus musculus molossinus DNA, clone:MSMG01-481P17.TJ, genomic survey sequence.

ACCESSION AG561783

VERSION AG561783.1 GI:48322481

KEYWORDS 1 SS.

SOURCE Mus musculus molossinus

ORGANISM Mus musculus molossinus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y. BAC end Sequences of Library MSMG01

2 (bases 1 to 776)

3 Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y. Direct Submision

4 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenho-chou Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

5 Tel:81-45-503-9111, Fax:81-45-503-9170

6 (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,

7 Clones are derived from the mouse BAC library MSMG01. For BAC library availability, please contact Kunya Abe (abe@rtc.riken.jp).

8 Tsukuba Institute, Bio Resource Center,

9 The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyada, Tsukuba, 305-0074 Japan

10 phone: 81-298-36-9189, fax: 81-298-36-9199

11 e-mail: abe@rtc.riken.jp

12 PRIMERS

13 Sequencing : TJ

14 LIBRARY

15 Vector : pBAC3.6

16 R.Site 1 : EcorI

17 R.Site 2 : EcorI.

18 Location/Qualifiers

19 1. 776

20 /organism="Mus musculus molossinus"

21 /mol_type="genomic DNA"

22 /sub_species="molossinus"

23 /db_xref="taxon:57486"

24 /clone="MSMG01-481P17.TJ"

25 /sex="male"

26 /issue_type="mixture of kidney and spleen"

27 /clone_lib="MSMG01 Mouse Male BAC library"

ORIGIN

Query Match 4.0%; Score 61.8; DB 9; Length 776;

Best Local Similarity 61.8%; Pred. No. 0.0035;

Matches 118; Conservative 0; Mismatches 67; Indels 6; Gaps 1;

QY 731 GGAGACCCCGCTCCCGGGGCTTAGCTGCGGGAGAGAGAGTAAAGTCCACTGG 790

DB 62 GGAATTCCAGATCTCTGGGAGCTGAGCTTCAAGAAAACAGGGGCCGCTCCAG 121

QY 791 GGAACCCCTAGACCGGATGACACGACTTTGTCTCAAGACACTGCGGCTCCGCTTC 850

DB 122 GAGGCCCAAGCAGCTGTATCTGCTGTCAATATGTGAACCTGCGCCCAACACC 181

QY 851 TGCAGCCCGCTCCAGAGTGTGACTTCCTCGGCTTCGACCCCTTACTGTGTGGGTC 910

DB 182 TGCAGCTCTCTG-----CTGAGCCTTCAGACATGTGCTCTGAGTCTCGGAC 235

QY 911 TGAGGGTCAGC 921

DB 236 TGAGGGGACG 246

RESULT 56

CE173623 262 bp DNA linear GSS 25-SEP-2003

LOCUS CE173623

DEFINITION tigr-gss-dog-17000326704476 Dog Library Canis familiaris genomic,

genomic survey sequence.

ACCESSION CE173623

VERSION CE173623.1 GI:35316113

KEYWORDS GSS.

SOURCE Canis familiaris (dog)

ORGANISM Canis familiaris

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

1 (bases 1 to 262)

2 Kirnesh, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Ruch, D.B., Delcher, A.L., Pop, M., Wang, W., Frazer, C.M. and Venter, J.C.

3 The dog genome: survey sequencing and comparative analysis

4 Science 301 (5641), 1898-1903 (2003)

5 22875432

6 14512627

7 Contact: Kirnesh EF

8 The Institute for Genomic Research

9 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,

10 Rockville, MD 20850, USA

11 Tel: 301-838-0200

12 Fax: 301-838-0208

13 Email: ekirnes@tigr.org

14 Class: Shotgun.

15 Location/Qualifiers

16 1. 262

17 /organism="Canis familiaris"

18 /mol_type="genomic DNA"

19 /strain="Standard Poodle"

20 /db_xref="taxon:9615"

21 /clone_lib="Dog Library"

22 /note="Site 1: BseXI; Libraries were prepared from peripheral blood"

ORIGIN

Query Match 4.0%; Score 61.6; DB 9; Length 262;

Best Local Similarity 66.7%; Pred. No. 0.004;

Matches 88; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 533 GGACGAGTGCCTCGGTGTGTCGGAGCGGGAGGCTCATCTGTGACGGCTGCG 592

DB 110 GAGTCTTCCCGCGGTGTGTCGAGACGCGCGAGAGCTCTCTGTCGACGCTGCCCTC 169

QY 593 GGCCTTCCACCTGAGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGAGAG 652

DB 170 CTCCTACACCTGCACTGCTTCAACCCGCGCTGCCGAGATCCCAAGGATGCT 229

QY 653 GTGCTCACTG 664

DB 230 CTGCCCGCGCTG 241

RESULT 57

BE234617

LOCUS BE234617 434 bp mRNA linear EST 10-JUL-2000

DEFINITION 141780 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.

ACCESSION BE234617

VERSION BE234617.1 GI:9019335

KEYWORDS EST.

SOURCE Sus scrofa (pig)

ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.

1 (bases 1 to 434)

2 Fahrenkrug, S.C., Smith, T.P.L., Frøking, B.A., Cho, J., White, J., Vallet, J., Wise, T., Rohrer, G.A., Pereira, G., Sultana, R., Quackenbush, J. and Keefe, J.W.

3 Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly

4 Mamm. Genome 13 (8), 475-478 (2002)

5 22213729

6 12226715

7 Contact: Smith TPL

8 USDA, ARS, US Meat Animal Research Center

9 PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGCAGCAGC
Plate: 79 row: P column: 20
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers

FEATURES

source

1..434
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_11b="MARC 4P1G"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

ORIGIN

Query Match 4.0%; Score 61.4; DB 2; Length 434;
Best Local Similarity 57.6%; Pred. No. 0.004;
Matches 110; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

528 AATGAGACGAGTGTGCGGTGTCGGAACGGCGGGAGCTCATCTGCTGTGACGGCTGC 587
|||||
239 AATGAGACGAGTGTGCGGTGTCGGAACGGCGGGAGCTCATCTGCTGTGAGAACTGT 298
|||||
588 CCTCGGGCCTTCCAGCTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGAGCC 647
|||||
239 CCCAAAGATTTCACCTGCTGTGTGTCACGTGCCACCGCTGGCAAAATTTCCGAGTGGAG 358
|||||
648 TGGAGGTGCTCAGCTGCTGCGAGCAACATTCACAGAGTGCAGCCCGGGCAGAGGAG 707
|||||
359 TGGATTGACCTCTTCCGAGACTTATCCAAACGAAAGTGATGATGATGATGCT 418
|||||
708 CCCCGGCCCA 718
|||||
419 CCCAGTCACAA 429

RESULT 58
CF792818/c 670 bp mRNA linear EST 21-OCT-2003
LOCUS 885797 MARC 4P1G Sus scrofa cDNA 3', mRNA sequence.
DEFINITION CF792818
ACCESSION CF792818
VERSION CF792818.1 GI:37797379
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE 1 (bases 1 to 670)
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
Noneman,D.J., Wray,J.E. and Keeler,J.W.
Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages

Unpublished (2003)
JOURNAL CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

COMMENT

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: TMM8013 row: I column: 11
Seq primer: TAGAAGCACAGTCGAGC.

FEATURES

source

Location/Qualifiers
1..670
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_11b="MARC 4P1G"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."

ORIGIN

Query Match 4.0%; Score 61.4; DB 7; Length 670;
Best Local Similarity 57.6%; Pred. No. 0.0042;
Matches 110; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

528 AATGAGACGAGTGTGCGGTGTCGGAACGGCGGGAGCTCATCTGCTGTGACGGCTGC 587
|||||
641 AATGAGACGAGTGTGCGGTGTCGGAACGGCGGGAGCTCATCTGCTGTGAGAACTGT 582
|||||
588 CCTCGGGCCTTCCAGCTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGAGCC 647
|||||
581 CCCAAAGATTTCACCTGCTGTGTGTCACGTGCCACCGCTGGCAAAATTTCCGAGTGGAG 522
|||||
648 TGGAGGTGCTCAGCTGCTGCGAGCAACATTCACAGAGTGCAGCCCGGGCAGAGGAG 707
|||||
521 TGGATTGACCTCTTCCGAGACTTATCCAAACGAAAGTGATGATGATGATGCT 462
|||||
708 CCCCGGCCCA 718
|||||
461 CCCAGTCACAA 451

RESULT 59
CF794185 672 bp mRNA linear EST 21-OCT-2003
LOCUS 889669 MARC 4P1G Sus scrofa cDNA 5', mRNA sequence.
DEFINITION CF794185
ACCESSION CF794185
VERSION CF794185.1 GI:37798748
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE 1 (bases 1 to 672)
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
Noneman,D.J., Wray,J.E. and Keeler,J.W.

Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages
Unpublished (2003)
JOURNAL CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.

Plate: TMM8013 row: I column: 11
Seq primer: GTAATACGACTCACTATGAGG.
Location/Qualifiers
1..672
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_11b="MARC 4P1G"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."

FEATURES

source

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30674548"
/issue_type="Upper Head"
/dev_stage="embryo 9.5 - 10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_1ib="NIH BMAP HS0"
/note="Organ: Upper Head; Vector: pYX-Asc; Site_1: Ecor I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGAAGTCAAT. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

ORIGIN

Query Match 3.9%; Score 60.8; DB 7; Length 804;
Best Local Similarity 69.2%; Pred. No. 0.0059;
Matches 83; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 545 CGTGTGTCGGAGCGGCGGAGCTCATCTGCTGACGCGCTGCCCTCGGACCTTCCACCT 604
DB 1 CGTCTCAAGACGCGGAGAGCTCTGCTGTGACACATGCCCTTCTTCCACCAT 60
QY 605 GCGCTGCTGCTCCCTCCGCTCGGAGATCCCATGTGACCTGAGAGTCTCCAGCTG 664
DB 61 CCACTGCTGAAACCCCGCTGCAAGATCCAAAGCGGAGATGCTGTGCTCCGCTG 120

RESULT 64

CNS006XK/c

LOCUS 935 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #

BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

AL066051.1 GI:4945019

VERSION

GSS.

Drosophila melanogaster (fruit fly)

KEYWORDS

Drosophila melanogaster

SOURCE

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 935)

REFERENCE

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazuo Oosawa and

Aaron Hammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain Y2; on bw sp, the same strain used for the BDGP's

p1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

Location/Qualifiers
1..935/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

/clone="BACR14N09"

/clone_1ib="RPCI-98"

/note="end : T7"

ORIGIN

Query Match 3.9%; Score 60.4; DB 9; Length 935;
Best Local Similarity 31.0%; Pred. No. 0.0074;
Matches 122; Conservative 95; Mismatches 174; Indels 3; Gaps 1;

QY 574 GCTGTAGCGGTGCGCTCCCTCGGCGCTTCCACTGACCTGCTGCTCCCTCGGCGGAGA 633
DB 926 GGGSSSSG 867
QY 634 TCCCCAG--TGGACCTGAGAGTGTCTCCAGCTGCTGCGAGGCAACATCCAGAGTGC 690
DB 866 GCGCGGSG 807
QY 691 AGCCCCGAG 750
DB 806 SCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 747
QY 751 GCGTTAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 810
DB 746 SSG 687
QY 811 ACACGACTTGTGTACACAGACACTGCGCGCTCCGCTTCTGACGCCCGCTGCGAGTC 870
DB 686 SCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 627
QY 871 TGAATCTTCTGCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 930
DB 626 CG 567
QY 931 CTCCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 964
DB 566 CGGCGAGSGGAGSGAGSGAGSGAGSGAGSGAGSGAGSGAGSGAGSGAGSGAGSG 533

RESULT 65

BB814801/c

LOCUS

BB814801.1

DEFINITION

QV4-BN0090-020600-236-g06 BN0090 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BB814801.1 GI:10247035

VERSION

EST.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 359)

REFERENCE

Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Brijnes, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Coes, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,

O'Hare, M.J., Soares, F., Brentani, R.R., Reis, U.F., de Souza, S.J. and

Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

PUBMED

10737800

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=QV4-BN0090-020600-236-g06&cl=2000-06-02&cl=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 359.

FEATURES

source

Location/Qualifiers

```

1..359
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="adult"
/clone_lib="BN0090"
/note="Organ: breast_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

```

ORIGIN

Query Match 3.9%; Score 60; DB 2; Length 359;
 Best Local Similarity 61.5%; Pred. No. 0.0082;
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

QY 532 AGGACGAGTGTGCGGTGTGCGGAGCGCGGAGGAGCTCATCTGTCAGCGCTGCCCTC 591
    |||||
DB 187 AGGACTATTGCGAGGTGTGCGAGCAAGCGGTGATCATCTGTGTGATTAACCTGTCC 128
    |||||
QY 592 GGGCTTCCACCTGCGCTGCTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGACCTGGA 651
    |||||
DB 127 GTGCTTACCAACAGTGTGCTGCTGATCCGACATGAGAGGCTCCCGAGGCGCAAGTGA 68
    |||||
QY 652 GGTGCTCCAGCTGCTGCGAGCAACAGTCCAGAGG 687
    |||||
DB 67 GCTGCCACACTGCGAGAGGAGGCAATCCAGTGG 32
    |||||

```

RESULT 66
 CF137810 461 bp mRNA linear EST 09-SEP-2003
 LOCUS UI-HF-BN0-anh-f-11-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone
 DEFINITION IMAGE:3093236 5', mRNA sequence.
 ACCESSION CF137810.1 GI:3253254
 VERSION CF137810.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 461)
 REFERENCE Donald, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newcom Road, 4156 MEBR, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Louis Staudt
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/humanfl.html>
 Seq primer: pyx-5.

FEATURES

source

Location/Qualifiers

```

1..461
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3093236"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LT1)"
/clone_lib="NIH_MGC_50"
/note="Vector: pRT3-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (3.5-4.4kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

```

ORIGIN

Query Match 3.9%; Score 60; DB 6; Length 461;
 Best Local Similarity 61.5%; Pred. No. 0.0084;
 Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

QY 532 AGGACGAGTGTGCGGTGTGCGGAGCGCGGAGGAGCTCATCTGTCAGCGCTGCCCTC 591
    |||||
DB 122 AGGACTATTGCGAGGTGTGCGAGCAAGCGGTGATCATCTGTGTGATTAACCTGTCC 181
    |||||
QY 592 GGGCTTCCACCTGCGCTGCTGCTCCCTCCGCTCCGAGAGATCCCAAGTGGACCTGGA 651
    |||||
DB 182 GTGCTTACCAACAGTGTGCTGCTGATCCGACATGAGAGGCTCCCGAGGCGCAAGTGA 241
    |||||
QY 652 GGTGCTCCAGCTGCTGCGAGCAACAGTCCAGAGG 687
    |||||
DB 242 GCTGCCACACTGCGAGAGGAGGCAATCCAGTGG 277
    |||||

```

RESULT 67
 BU429642 544 bp mRNA linear EST 09-SEP-2002
 LOCUS UI-HF-BN0-ank-f-09-0-UI.r2 NIH_MGC_50 Homo sapiens cDNA clone
 DEFINITION IMAGE:3064481 5', mRNA sequence.
 ACCESSION BU429642
 VERSION BU429642.1 GI:22768129
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 544)
 REFERENCE NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 CDNA Library Preparation: M.B. Soares Lab
 CDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://bio.llnl.gov/bbrp/image/image.html>
 Seq primer: M13 Forward
 Location/Qualifiers
 1..544
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:3064481"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LT1)"
 /clone_lib="NIH_MGC_50"

QY 652 GGTGCTCAGCTGCTGCTGAGGCAACGTCCAGAGG 687
EST.
Homo sapiens (human)
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgc.nci.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: LMC1265 row: h column: 20
High quality sequence stop: 650.
Location/Qualifiers
1. 659
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4558915"
/issue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 3.9%; Score 60; DB 4; Length 699;
Best Local Similarity 61.5%; Pred. No. 0.0088;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTCGCGGAGGAGGAGCTCTGCTGACGCTGCTCC 591
DB 35 AGGACTATTGCGAGGTGTCACAGAGCGGTGATCATCTGTGTATCACTGTCC 94
QY 592 GGGCCCTTCACCTGCGCTGCTGCTCCCTCCGCTCCGAGAGATCCCACTGGAGCTTGA 651
DB 95 GTGCTTACCAACATGCTGCTGCTGATCCGACATGAGAAAGCTCCCGAGGCGCAACTGA 154
QY 652 GGTGCTCAGCTGCTGCGAGGCAACGTCCAGAGG 687
DB 155 GGTGCCACACTGCGAGAAAGGACATCCAGTGG 190

RESULT 71
BE258675 700 bp mRNA linear EST 27-FEB-2001
LOCUS BE258675
DEFINITION 60111634F1 NIH_MGC_16 Homo sapiens CDNA clone IMAGE:3156962 5',
mRNA sequence.
ACCESSION BE258675

VERSION BE258675.1 GI:9129168
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgc.nci.nih.gov/
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: LMC163 row: g column: 11.
Location/Qualifiers
1. 700
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3156962"
/issue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC 16"
/note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 3.9%; Score 60; DB 2; Length 700;
Best Local Similarity 61.5%; Pred. No. 0.0088;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTCGCGGAGGAGGAGCTCTGCTGACGCTGCTCC 591
DB 70 AGGACTATTGCGAGGTGTCACAGAGCGGTGATCATCTGTGTATCACTGTCC 129
QY 592 GGGCCCTTCACCTGCGCTGCTGCTCCCTCCGCTCCGAGAGATCCCACTGGAGCTTGA 651
DB 130 GTGCTTACCAACATGCTGCTGCTGATCCGACATGAGAAAGCTCCCGAGGCGCAACTGA 189
QY 652 GGTGCTCAGCTGCTGCGAGGCAACGTCCAGAGG 687
DB 190 GGTGCCACACTGCGAGAAAGGACATCCAGTGG 225

RESULT 72
BE176589 731 bp mRNA linear EST 21-JUN-2000
LOCUS BE176589
DEFINITION RC3-HT0585-010400-023-c04 HT0585 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE176589
VERSION BE176589.1 GI:8639318
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS
Diaz Neco E., Garcia Correa R., Verjovski-Almeida S., Brito M.R.,
Nagai M.A., da Silva M. Jr., Zago M.A., Bordin S., Costa F.F.,
Goldman G.H., Carvalho A.F., Matsushima A., Bala G.S., Simpson D.H.,
Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V.,
O'Hare M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and
Simpson A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
JOURNAL sequence tags
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PABSP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?fl=&tz=RC3-HT0585-010
400-023-c04&tz=2000-04-01&tz=1)
Seq primer: puc 18 forward
High quality sequence start: 22
High quality sequence stop: 719.
Location/Qualifiers
1. 731
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="HT0585"
/dev_stage="Adult"
/note="Organ: head, neck; Vector: puc18; Site: 1: Sma1;
Site: 2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Query Match 3.9%; Score 60; DB 2; Length 731;
Best Local Similarity 61.5%; Pred. No. 0.0089;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTGCGGTGTGCGGACGCGGAGGCTCATCTGTGACGGCTGCCCTC 591
|||||
DB 376 AGGACTATTGCGAGGTGTGCGACGAAGCGGTGAGATCATCTGTGTATCACTGTCCCC 425
|||||
QY 592 GGGCCTTCCACCTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCGACATGGAACCTGGA 651
|||||
DB 436 GTGCTTACCACTGCTGCTGCTGATCCGACATGAGAGGCTCCCGAGGCGCAAGTGA 495
|||||
QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGG 687
|||||
DB 496 GCTGCCACACTGCGAAGAGGAGCATCCATGCGG 531
|||||

RESULT 73 750 bp mRNA linear EST 20-JUL-2000
BE297517
LOCUS BE297517
DEFINITION 601178063F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3533391 5',
mRNA sequence.
ACCESSION BE297517
VERSION BE297517.1 GI:9181002
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://img.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory

FEATURES
source
1. 750
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="IMAGE:353391"
/tissue_type="rhabdomyosarcoma"
/lab_host="PH10B (phage-resistant)"
/clone_lib="NIH_MGC_17"
/note="Organ: muscle; Vector: pOTB7; Site: 1: EcoRI;
Site: 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN
Query Match 3.9%; Score 60; DB 2; Length 750;
Best Local Similarity 61.5%; Pred. No. 0.0089;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 532 AGGACGAGTGTGCGGTGTGCGGACGCGGAGGCTCATCTGTGACGGCTGCCCTC 591
|||||
DB 6 AGGACTATTGCGAGGTGTGCGACGAAGCGGTGAGATCATCTGTGTATCACTGTCCCC 65
|||||
QY 592 GGGCCTTCCACCTGCGCTGCTGCTCCCTCCGCTCCGGAGATCCCGACATGGAACCTGGA 651
|||||
DB 66 GTGCTTACCACTGCTGCTGCTGCTGATCCCGACATGAGAGGCTCCCGAGGCGCAACTGGA 125
|||||
QY 652 GGTGCTCCAGCTGCTGCGACGCAACAGTCCAGAGG 687
|||||
DB 126 GCTGCCACACTGCGAAGAGGAGCATCCATGCGG 161
|||||

RESULT 74 817 bp mRNA linear EST 13-JUL-2000
BE259148
LOCUS BE259148
DEFINITION 601108075F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3344251 5',
mRNA sequence.
ACCESSION BE259148
VERSION BE259148.1 GI:9129658
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://img.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM130 row: e column: 20.
Location/Qualifiers
1. 817
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3344251"

QY 652 GGGCTCCAGTGCCTGCGAGGACGACAGTCCAGGAGG 687
| | | | | | | | | | | | | | | | | | | | | |
Db 124 GCTGCCACACTGCGAGAGGAGGACATCCAGTGGG 159
| | | | | | | | | | | | | | | | | | | | | |

RESULT 77
LOCUS BG394104
DEFINITION 602456272f1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4579048 5',
mRNA sequence.
ACCESSION BG394104
VERSION BG394104
KEYWORDS GI:13287552
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L16M1295 row: 0 column: 17
High quality sequence stop: 635.
Location/Qualifiers
1. .974
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4579048"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_11b="NIH_MGC_16"
/note="Organ: eye; Vector: pOT7; Site: 1: XhoI; Site: 2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 3.9%; Score 60; DB 4; Length 974;
Best Local Similarity 61.5%; Pred. No. 0.0091;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGACGGGAGGAGTCACTGCTGTGACGGCTGCCCTC 591
| | | | | | | | | | | | | | | | | | | | | |
Db 35 AGGACTATTGCGAGGTGTGCGACGAGGGGTGATCATCTGTGTACTGTGCCCC 94
| | | | | | | | | | | | | | | | | | | | | |

QY 592 GGGCTTCCACTGCGCTGCTGCTCCCTTCCTCGCGGAGATCCCAAGTGGAGCTGGA 651
| | | | | | | | | | | | | | | | | | | | | |
Db 95 GTGCTTACCACTGCTGCTGCTGATCCGACATGGAAGGCTCCCGAGGGCAAGTGA 154
| | | | | | | | | | | | | | | | | | | | | |

QY 652 GGGCTCCAGTGCCTGCGAGGACGACAGTCCAGGAGG 687
| | | | | | | | | | | | | | | | | | | | | |
Db 155 GCTGCCACACTGCGAGAGGAGGACATCCAGTGGG 190
| | | | | | | | | | | | | | | | | | | | | |

RESULT 78
LOCUS BM457082
DEFINITION 989 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6406806 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583326
5', mRNA sequence.
ACCESSION BM457082

VERSION BM457082.1 GI:18506122
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L16M12346 row: d column: 15
High quality sequence stop: 658.
Location/Qualifiers
1. .989
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5583326"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_11b="NIH_MGC_92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 3.9%; Score 60; DB 4; Length 989;
Best Local Similarity 61.5%; Pred. No. 0.0091;
Matches 96; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGACGGGAGGAGTCACTGCTGTGACGGCTGCCCTC 591
| | | | | | | | | | | | | | | | | | | | | |
Db 38 AGGACTATTGCGAGGTGTGCGACGAGGGGTGATCATCTGTGTACTGTGCCCC 97
| | | | | | | | | | | | | | | | | | | | | |

QY 592 GGGCTTCCACTGCGCTGCTGCTCCCTTCCTCGCGGAGATCCCAAGTGGAGCTGGA 651
| | | | | | | | | | | | | | | | | | | | | |
Db 98 GTGCTTACCACTGCTGCTGCTGATCCGACATGGAAGGCTCCCGAGGGCAAGTGA 157
| | | | | | | | | | | | | | | | | | | | | |

QY 652 GGGCTCCAGTGCCTGCGAGGACGACAGTCCAGGAGG 687
| | | | | | | | | | | | | | | | | | | | | |
Db 158 GCTGCCACACTGCGAGAGGAGGACATCCAGTGGG 193
| | | | | | | | | | | | | | | | | | | | | |

RESULT 79
LOCUS BF309509
DEFINITION 1011 bp mRNA linear EST 21-NOV-2000
601892016r1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4137843 5',
mRNA sequence.
ACCESSION BF309509
VERSION BF309509
KEYWORDS GI:11256902
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC


```

/dev strage="Adult"
/clone.lib="EN0063"
/notes="Organ: lung normal; Vector: puc18; Site_1: Smal;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORSTS PCR (U.S. Veterans Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

ORIGIN

Query Match	3.9%	Score 59.8;	DB 2;	Length 426;
Best Local Similarity	58.9%	Pred. No. 0.0093;		
Matches 103; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0;

Qy	512	AGGAGGAGGTGC	CCGATGTTCGGGAA	ACGGGGGAGCTACT	CTGTGTGACGGGTGCCCTC	591
Db	247	AGGATTACTGTAG	GTGTGCCAGACGGGTGGGAA	TATTCTGTGTGACCACTG	CGCCCTC	306
Qy	552	GGGCTTTCACCT	TGGCTGCTGCTCCCTTC	CGTCCGGGAGATCC	CACTGTGGAACTCTGGA	651
Db	307	GTGCTTACCACT	CGTGTGCTGTGATCT	GTAGGTTGACCGGGCT	CCAGAGGGCAAAATGGA	366
Qy	652	GGTGTCCAGCT	GCCTGACAGCAACAT	GCACGAGGTGACGCC	CCGGGACAGAGA	706
Db	367	GCTGCTCTCAT	CTGTGAGGAAGGGGCT	CAGTGGAGGCCA	AGAGGAAGAAAGA	421

RESULT	82
BUI93516	
LOCUS	BUI93516
DEFINITION	BUI93516 885 bp mRNA
	AGENCECOURT_7840454 NIH_MGC_67 Homo sapiens CDNA clone IMAGE:614630
	5', mRNA sequence.

ACCESSION	BU193516
VERSION	BU193516.1
KEYWORDS	GI:22707500
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 885)	NIH-MGC	http://mgc.ncbi.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
	Contact: Robert Strusberg, Ph.D.			

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13474 row: k column: 07
High quality sequence stop: 724.

FEATURES

source

Location/Qualifiers
1. .885

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6146430"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_1b="NH1_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."

```

ORIGIN

Query Match	3.9%	Score 59.8;	DB 5;	Length 885;
Best Local Similarity	58.9%;	Pred. No. 0.01;		
Matches 103; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0;

QY	522	AGAGCAGATGTCCCTGTGTCCGGAGCGGGGGAGCTACTCTGCTGTAGAAGCTGTGCCCTC	521
Db	333	AGGATTTACTGTGAGGATGTGCCACGACAGGGGGGGAATTTATCTGTGTGACACTGTGCCCTC	442
QY	532	GGGCCTTTCACCTGTGACTCTGTCCCTGTCCCGCTCCGGGAGATCCCCAGTGTGGACCTTGA	651
Db	443	GTGCTTACCACTCTCGTCTGTGCTTGTATCTCGACTTGAACCGGGCTCCAGAGGGCGAATTTGA	502

Qy 652 GGTGCTCCAAGCTGCCTGCAAGCACAAGTCCAGAGAAGTGACGCCCGGGCAGAAGGA 706
| | | | | | | | | | | | | | |
Db 503 GCTGCCCTCACTGTGAGAAGAGAGGGGTCCAGTCGGAGGCCAAGAGGAAGAAGA 557

RESULT 83				
BU527069				
LOCUS	BU527069	918 bp	mRNA	linear
DEFINITION	AGENCOURT_10155962 NIH_MGC_101 Homo sapiens cDNA clone			EST 13-SEP-2002
	IMAGE:6536820 5', mRNA sequence.			

ACCESSION	BU527069	GI:22837510
VERSION	BU527069.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE
1 (bases 1 to 918)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 918)	NIH-MGC	http://mgc.nci.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)	Contact: Robert Strausberg, Ph.D.	

Email: cgapbs-i@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LUCM2698 row: m column: 12
High quality sequence stop: 666.

FEATURES

source

Location/Qualifiers
1. .918

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="6536820"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH MGC 101"
/note="Organ: Lung, Vector: pOTB7, Site_1: EcoRI, Site_2: XhoI, cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."

```

ORIGIN

Query Match	3.9%	Score 59.8;	DB 5;	Length 918;
Best Local Similarity	58.9%;	Pred. No. 0.01;		
Matches 103; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0

QY	532	AGACGAGTGTGCCCTGTGTGTCCGACCGCGGGAGCTCATCTGCTGTGACCGGCTGCCTC	591
Db	145	AGGATTACTGTGAGGTGTGTCCGACAGGGTGGGAAATATTCTGTGTGACACTTGCTTC	204

QY	592	GGGCGCTTCCACTGAGCTGCTGTCCCTCCGCTCCGGGAAATCCCAAGTGGGACCTGGA	651
Db	205	GTGCTTACCACTCGTGTGCTTGTATCTCGAGCTTGAACCGGAGCTCCAGAGGGCAATGGA	264

652 GGTCCTCCAGCTGCTGCAGGCAACAGTCCAGAGAGTGACCCCGGCGAGAGA 706

Db 265 GCTGCTCACTGTGAGAGAGGGGCTCCAGTGGAGGCCAAGAGGAAGA 319

RESULT 84
LOCUS BUS28543
DEFINITION BUS28543 943 bp mRNA linear EST 13-SEP-2002
IMAGE:6538496 5', mRNA sequence.

ACCESSION BUS28543
VERSION BUS28543
KEYWORDS BUS28543.1 GI:22838984
SOURCE EST.
ORGANISM Homo sapiens (human)
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/1 (bases 1 to 943)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LCM2703 row: C column: 08
High quality sequence start: 24
High quality sequence stop: 683.
Location/Qualifiers
1. 943
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6538496"
/issue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pOT7; Site 1: EcoRI; Site 2: XhoI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

ORIGIN
Query Match 3.9%; Score 59.8; DB 5; Length 943;
Best Local Similarity 58.9%; Pred. No. 0.01;
Matches 103; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGGTGTGCGGACGCGGGAGCTCATCTGTGACGCTGCCCTC 591
DB 178 AGGATTACTGTGTGCGGTGTGCGGACGCGGGAGAAATTATCTGTGTGACACTGCCCTC 237
QY 592 GGGGCTTCACCGGCTGCTGCTGCCCTGCCCTCCGGAGATCCCAAGTGGAGACCTGGA 651
DB 238 GTGCTTACCACTCGTGTGCTGTGATCTGAGCTTGAACCGGGCTCCAGAGGGCAAAATGA 297
QY 652 GGTGTCTCAGCTGCTGCAAGCAAGTCCAGAGAGGTGACGCCCGGCGAGAGGA 706
DB 298 GCTGCTCACTGTGAGAGAGGGGTCCAGTGGAGGCCAAGAGGAAGA 352

RESULT 85
LOCUS B0722521/c
DEFINITION B0722521 1104 bp mRNA linear EST 16-JUL-2002
AGENCOURT 8219490 luepki sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6188186 5', mRNA sequence.

ACCESSION B0722521
VERSION B0722521.1 GI:21861418
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/1 (bases 1 to 1104)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James R. Luepki
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM13583 row: G column: 03
High quality sequence stop: 297.
Location/Qualifiers
1. 1104
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6188186"
/sex="male"
/issue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_lib="luepki_sympathetic_trunk"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site 1: NotI; Site 2: SalI; CDNA made by oligo-dT priming. Directionally cloned using the following adaptor: 5'-TCGACCCAGACCTCCG-3' and 5'-GACTAGTCTTGAATCGGAGCGGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Luepki, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."

ORIGIN
Query Match 3.8%; Score 59.4; DB 5; Length 1104;
Best Local Similarity 41.8%; Pred. No. 0.013;
Matches 270; Conservative 0; Mismatches 374; Indels 2; Gaps 1;

QY 712 GGGCCAGAGCCAGCCGTGAGAGACCCGCTCCCGGAGCTTGAAGTGGCGGAGAG 771
DB 912 GNCGCCCGGAGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853
QY 772 AGTGAAGATCCACTGTGGGAACCCCTAGCCGCGCATGACAGACTCTTGTACAGC 831
DB 852 CCG 793
QY 832 ACTGCGCGCTCGCTTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 891
DB 792 CCG 733
QY 892 CCTACTGTGTGTGTGCTGTGAGAGGTGACAGAACTGCTCTGTGTGCGCTGGG 951
DB 732 GCG 673
QY 952 TGGCGGAGATGTACGAGCGTGTGCGGTGTACTCATGCGCGCGCTTCACTGAC 1011
DB 672 GGGGCG 615
QY 1012 GCTGCACTTCCAGCGCGCACTCCCGGCGCGGAGCGGAGCTTGCGCTGAGATCTGCT 1071
DB 614 GCGGCG 555

[illegible]

RESULT	86
B0672163	
LOCUS	B0672163 1132 bp mRNA linear EST 15-JUL-2002
DEFINITION	AGENCOURT_8302070 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6274983 5', mRNA sequence.

VERSION	BQ672163.1	GI:21782997
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE	1 (bases 1 to 1132)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

CNA Library Preparation: Rubin Laboratory
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MCC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
<http://image.llnl.gov>
plate: LILCM2457 row: g column: 16
High quality sequence stop:306.

```

FEATURES
source
Location/Qualifiers
1..1132
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6274983"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH MGC_102"
/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."

```

Query Match	3.8%	Score 58.8;	DB 5;	Length 1132;
Best Local Similarity	31.8%;	Pred. No. 0.017;		
Matches 276;	Conservative	0;	Mismatches 592;	Indels 0;
			Gaps	0;

Db	224	GCCTGGGCGCAGGGCAGAGACCGGCTTTCTCTGCGGGAGCGGGGGGGGCTCTGGAGC	283
Oy	98	CTGCTATATCCAGAGGTCAAAGATCCACTGGGATGCTCATCTTTCTGTCACG	157
Db	284	CGCCCGGGGGGTGGGTNN	343
Oy	158	CATGTTTCTTAATGGGGTAGAAGCAGGTGGAGAGACTCCTGGGCTGGCCACT	217
Db	344	NN	403
Oy	218	GCCCTGTGAAGAGGTTCACTGTGTTGTGTACAGTTCCGGGGCCCTGAAACGACGA	277
Db	404	NN	463
Oy	278	GCTCTCAAGAAACCGGTTTTCTTCCCAATAGGAGTGGCCCCGGGGGTGTCTGTGAG	337
Db	464	NNNNNGNN	523
Oy	338	ACCAGATGATGAGGGAACAGGTGCTCAGGGCAGATTTACGGCCCTTGACATAGGAGC	397
Db	524	GG	583
Oy	398	AGGCGAAGACTGGGGAGTTTCAAGTACCAGAGATGCTGCTGGGGGAGCTGTTTGGGAA	457
Db	584	NNGG	643
Oy	458	GGAGGTGGCTCTCAGAGGGGTGTGCAACCCAGCCAGTCTGCAATGGAGCTCTTGTACT	517
Db	644	GGGGGGGGGAGNNNNGG	703
Oy	518	GTGCCAAGAAATAGAGACGAGTGTGCTGTGTGTGGGACGGCGGGAGCTCATCTGTG	577
Db	704	NNGGGGGGGGGGGGGGGGGGGGNNNNCGGAGGGGGGGGGGGGGGGGGGNNNGGGGGG	763
Oy	578	TGAAGGCTGCTCGGCTTTCCACTGAGCCTGTGCTGCTCCCTCCGCTCCGGAGATCC	637
Db	764	CCGGGGGGGGGGCGGCGGCCCCCCCCCGGGGGCGGCCCCCGGNNNGCGCGCGGGGCGC	823
Oy	638	CAGTGGGACTGAGAGTGTCAAGCTGCTTGCAGAGCAAGTCCAGAGAGGTGACCCCG	697
Db	824	GGCGGGGGGNNGGCCCCCGGGGGGTCCCGCGGGCGGGGCGCGGCGCGCGCGC	883
Oy	698	GGCAGAGAGGCCCCGGCCCCCAGAGACCACTGTGAAAGACCCCGTCCCGCCGGGGCTTAA	757
Db	884	GGGGCGCGCGCGCCCGCGCGCGCGGGGCGGACCCCGGGCGGCGCGCGCGCGCGC	943
Oy	758	GTCCGGCGGAGAGGATTAAGAGGTTCACCTGGGGAACCCCTTACCGGATGACAGACGAC	817
Db	944	GGCCCCCGCGACGCGCGCGCGCGCGCGGGAGGAGCGCGCGCGCGCGCGCGCGCGCGC	1003
Oy	818	TCTTGTCTAACAGACACTGCGCGGCTCCGCTTGTGACAGCCCGGCTGCCAGGTTGGAATC	877
Db	1004	CGCNCGCGCCCCCGGGCGGGCGGGCGCGCGCCCCCGGGCGGGCGCGGGCGCGCGCGC	1063
Oy	878	CTCGGCTCTGACCCCTCATCTGTGTGG	905
Db	1064	CGGGGCGGGCGCNCCTGGGTCCGGCGGGG	1091

RESULT	87				
BM450237					
LOCUS	BM450237	1350 bp	mRNA	linear	EST-05-PEB-2002
DEFINITION	AGENCOCURT_6393336 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528315				
ACCESSION	5', mRNA sequence.				
VERSION	BM450237				
KEYWORDS	BM450237.1 GI:18499277				
SOURCE	EST.				
ORGANISM	Homo sapiens (human)				
	Homo sapiens				
	Bukayolota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1350)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Straube, Ph.D.
 Email: c9apbs-x@mail.nih.gov
 Tissue Procurement: ATCC/CDT/DRP
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: LLM42204 row: s column: 12
 High quality sequence stop: 370.
FEATURES
 source
 1. 1350
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5528315"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NH MGC 72"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."
ORIGIN
 Query Match 3.8%; Score 58.6; DB 4; Length 1350;
 Best Local Similarity 41.0%; Pred. No. 0.02;
 Matches 267; Conservative 0; Mismatches 378; Indels 7; Gaps 1;
 QY 477 GTGCTGACCCCAAGCTGATGAGGCTCTTCTGTCAGAGAAATGAGAG 536
 DB 705 GGGCCCCCGGGCGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 764
 QY 537 GAGTGTGCGGTGTGTGGAGAGCGGGGAGAGCTATGTCTGTGA 596
 DB 765 CGGCG 824
 QY 597 TTCCAGCTGAGCTGCTGCTCCCTCCGCTCCGGGAGATCCCAAGTGAAGTGC 656
 DB 825 CG 884
 QY 657 TCCAGCTGCTGACGACCAAGTTCAGAGAGTGCAGCCCGGAGAGAGAGAG 716
 DB 885 CCG 937
 QY 717 CAGAGAGCAACCGGTGAAGCCCGCTCCCGCGCGCGCGCGCGCGCGCGCG 776
 DB 938 CCG 997
 QY 777 AGAGTTCACCTGAGGAAACCTTAGCGGATGAGACAGACTTTGTACAAGAC 836
 DB 998 GGGGGGCG 1057
 QY 837 CCGGCTCCGCTTGTGACAGCCCGCTGCAAGTGTGAGTCTCTGCGCCCTG 896
 DB 1058 NNCCNCCCG 1117
 QY 897 CTGTGTGTGGGTCTGAGGATGAGAGACTGTGCTGTGTGTGTGTGTGTGT 956
 DB 1118 NCCNCCG 1177
 QY 957 GGAGATGTATCGGACGTGTGCGGTGTACTCACTGCGCGCTGCTTCACTG 1016
 DB 1178 CCGGNNCCG 1237
 QY 1017 CACTTTCAGCGCGGACCTTCGCGCGCGGAGAGGAGCTGTGCTGATCTG 1076
 DB 1238 CCG 1297
 QY 1077 GAGGTGACCCCAAGCCCTGTGAGAGGGGTGTGTGTGTGTGTGTGTGT 1128

RESULT 88
LOCUS CB606244 416 bp mRNA linear EST 16-MAY-2003
DEFINITION AMGNNUC:SRPB2-00055-cl1-A srpb2 (10220) Rattus norvegicus cDNA
ACCESSION CB606244
VERSION CB606244.1 GI:29545856
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
REFERENCE
 1 (bases 1 to 416)
 Amgen EST Program.
AUTHORS Amgen EST Program
TITLE Amgen Rat EST Program
JOURNAL Unpublished (2003)
COMMENT Contact: Dan Fitzpatrick
 Amgen, Inc
 One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
 Tel: 805 447-4881
 Plate: 00055 row: c column: 11.
FEATURES
 source
 1. 416
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="srpb2-00055-cl1"
 /tissue_type="prostate tissue"
 /clone_1ib="srpb2 (10220)"
 /note="Vector: pSPORT1; Site: 1: SalI; Site 2: NotI; rat prostate normalized double selected poly(A+) mRNA size fraction > 1 kb"
ORIGIN
 Query Match 3.8%; Score 58.4; DB 6; Length 416;
 Best Local Similarity 60.9%; Pred. No. 0.019;
 Matches 95; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
 QY 532 AGGACGAGTGTGCGGTGTGTGCGAGCGCGGAGCTATCTGTGTGACGGCTG 591
 DB 14 AGGACTATATGAGAGGTGTGCTGACCAAGCGGAGATCATCTGTGTATAC 73
 QY 592 GGGCTTCACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 651
 DB 74 GAGCTACCATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 133
 QY 652 GGTGTCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 687
 DB 134 GCTGCCACACTGTGAGAGAGAGGTATCCATGTGG 169
RESULT 89
LOCUS CNS00720 932 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BAC114B09 of RP11-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION CNS00720
VERSION CNS00720.1 GI:4945205
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
 1 (bases 1 to 932)
 Genome.
AUTHORS Direct Submission

JOURNAL Submitted (02-JUN-1999) GENOSCOPE - Centre National de Sequencage
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see [http://www.flyfritzy.org/TheBDGP/Drosophila](http://www.flyfritzy.org/TheBDGP/Drosophila%20melanogaster%20BAC%20library)
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aarton Mammossier in Pieter de Jong's Laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPc1-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2: cn bw sp, the same strain used for the BDGP's
 pl and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 location/Qualifiers

FEATURES
 source
 1..932
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR14809"
 /clone_11b="RPc1-98"
 /note="end : 77"

ORIGIN

Query Match	3.8%	Score 58.4;	DB 9;	Length 932;
Best Local Similarity	31.0%;	Pred. No. 0.021;	Mismatches 162;	Indels 0;
Matches 109;	Conservative 81;			Gaps 0;
Qy	544	CCGTTGTGCGGAGACGCGCGGAGAGCTCATCTGCTGTGACGCGTGCCTTCGAGCTTCACCC	603	
Dp	861	SCGCGSSSSCGCCGCCGCGSSGGGCGCCSSGCGSSCGCCSSCGCCSSCGCCSSCGCCSSCGCCSS	802	
Qy	604	TGCGCTGCTCTGCTCCCTTCCTCCGTCGCGAGATCCCAATGAGGACCTGAGAGTGTCTCACT	663	
Dp	801	CSSCGCGSSSSCGSSCGCCGCGSSCGCCSSCGCGSSCGCGSSCGCCSSCGCCSSCGCGSSCGCG	742	
Qy	664	GCTTCAGAGCAACACTCCAGAGGTGACAGCCCCCGGACAGAGAGCGCCCGCCCAAGAGAGC	723	
Dp	741	SGGGSSGSGGCGSSGCGSSGCG	682	
Qy	724	CACCGTGAAGACCCCGCTCCCGCGCGGCTTAAAGTGTGCGCGGAGAGAGAGTAAAGAGTTC	783	
Dp	681	SGCGCGCGCGCCSCCGSSCG	622	
Qy	784	CACCTGGGGGAACCCCTAGCCCGGCAATGACACGACTCTTGTCATGAAGACACTGCGCGGCTC	843	
Dp	621	SCGCGCCCGCGSSCGSSGCGCCGCMCAATMAVAIVSVCCCGCCGMAASCCCGCTVSGGSC	562	
Qy	844	CGCCTTGTGACAGCCCGCTGCGCAGGCTGGAAGTCTCTCGAGCCCTGACCCCTT	895	
Dp	561	CSCGSCCGCCCGCGCAATCAAGCGCMAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	510	

RESULT 90
 BP761124
 BP761124 mouse (C57BL/6) pancreatic islet library with
 recombination-based method Mus musculus cDNA clone m1c01051 5',
 mRNA sequence.
 BP761124
 BP761124.1 GI:50219822
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (baes 1 to 302)
 Nishimura,M., Yokoi,N., Miki,T., Horikawa,Y., Yoshioka,H.,
 Takeda,J., Ohara,O. and Seino,S.

TITLE	Construction of a multi-functional cDNA library specific for mouse									
JOURNAL	Published (2004)									
COMMENT	Contact: Suenmu Seino Division of Cellular and Molecular Medicine Kobe University Graduate School of Medicine 7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan Tel: 81-78-382-5360 Fax: 81-78-382-5370 Email: seinumed.kobe-u.ac.jp.									
FEATURES	Location/Qualifiers									
SOURCE	1..302									
ORIGIN	/organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" /clone="mic01051" /sex="male" /tissue_type="pancreatic islet" /dev_stage="adult" /clone_lib="mouse (C57BL/6) pancreatic islet library with recombination-based method"									
Query Match	3.8%	Score 58.2;	DB 5;	Length 302;						
Best Local Similarity	58.3%	Pred. No. 0.021;								
Matches 102;	Conservative 0;	Mismatches 73;	Indels 0;	Gaps 0						
Qy	532	AGAGCAGATGTCGCTGTGTGCGGACGGCGGGGAGCTCATCTGCTGTACAGGCTGCCCCC	591							
Db	77	AGGATVACTGTGAGAGTGTGCCAGCAGGGCGGGGAATATTTCTGTGCGACACCTGCCCC	136							
Qy	592	GAGGCTTCCACCTGAGCTGCTGCTGCTCCCTCCGCTCCGGAGATVCCCAAGTGGACCTGGA	651							
Db	137	GTGCTACCACTGTGTGCTTGACCCCGAGCTTGACCGAGCTCTGAGGGCAAGTGA	196							
Qy	652	GGTGCTCCAGTGCCTTCGACGCAACAGTTCACAGAGTGTCAAGCCCCGGGCAGAGGA	706							
Db	197	GCTGCCCCCACTGTGAGAGGAGGGGGGTGAGTGGAGGCCAAGAGAGAGAGA	251							
RESULT 91										
CF535771	357 bp	mRNA	linear	EST 12-SRP-200						
LOCUS										
DEFINITION	U1-M-GH0-cha-c-05-0-U1.r1 NIH_BMAP_GH0 Mus musculus cDNA clone									
ACCESSION	IMAGE:30533716 5', mRNA sequence.									
VERSION	CF535771									
KEYWORDS	CF535771.1 GI:34587739									
SOURCE	EST.									
ORGANISM	Mus musculus (house mouse)									
REFERENCE	Mus musculus									
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.									
TITLE	1 (bases 1 to 357)									
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .									
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cga@bcrfemail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Distribution information can be found at http://genome.uiowa.edu/distribution/mouse1.html This clone was contributed by the Brain Molecular Anatomy Project (BMAP)									
FEATURES	The following repetitive elements were found in this cDNA sequence: 302-357, >(GGA)nhsimple_repeat Seq primer: PYX-5.									
SOURCE	Location/Qualifiers									
ORIGIN	1..357									
ORGANISM	"Mus musculus"									

MG1:661037
Seq primer: -28m3 rev2 ET from Amerham
High quality sequence stop: 419.
Location/Qualifiers

FEATURES
source

```
1..428
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1247349"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares_mammary_gland_NBMG"
/note="Organ: mammary gland; Vector: pT73D-Pac
(pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTTCACATCTGAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
```

ORIGIN

Query Match 3.8%; Score 58.2; DB 1; Length 428;
Best Local Similarity 58.3%; Pred. No. 0.021;
Matches 102; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
QY 532 AGGACGAGTGTCCCTGTGTCTGGAGCGGGGAGCTCATCTGTACGCTGCCCTC 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 AGGATTACTGTAGTGTGTCACAGCGGGGAAATTATCTGTGCACACCTGCCCC 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 592 GGGCCCTTCACCTGGCGCTGTCCCTCCGCTCCGGAGATCCCAAGTGGAGCTTGA 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 303 GTGCTTACCACTCTGCTGTGACCCGAGCTTACCAAGGCTCTTGAAGGCAATGGA 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 652 GGTGCTCCAGCTGCTGCAGGCAACAGTCCAGAGGTGACCCCGGCAGAGGA 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 GCTGCCACCACTGTGAGAGGAGGGGTGCAGTGGAGGCCAAGAGAGAGAGA 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 94
CBS20637 452 bp mRNA linear EST 09-JUL-2003
LOCUS UI-M-GIO-cef-1-03-0-UI.r1 NIH BMAP_GIO Mus musculus cDNA clone
DEFINITION IMAGE:6840436 5', mRNA sequence.
ACCESSION CBS20637
VERSION CBS20637.1 GI:29353992
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 452)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
The following repetitive elements were found in this cDNA
sequence: 385-450, >(GGA)n#simple_repeat

Seq primer: PYX-5.
Location/Qualifiers

FEATURES
source

```
1..452
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6840436"
/tissue_type="whole brain"
/dev_stage="embryo 13.5, 14.5, 16.5, 17.5 dpc"
/lab_host="NIH BMAP G10"
/clone_lib="NIH BMAP G10"
/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel, first strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to RNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemn Chin, Ph.D.,
program coordinator."
```

ORIGIN

Query Match 3.8%; Score 58.2; DB 6; Length 452;
Best Local Similarity 60.4%; Pred. No. 0.021;
Matches 96; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
QY 532 AGGACGAGTGTCCCTGTGTCTGGAGCGGGGAGCTCATCTGTACGCTGCCCTC 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 251 AGGATTACTGTAGTGTGTCACAGCGGGGAAATTATCTGTGCACACCTGCCCC 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 592 GGGCCCTTCACCTGGCGCTGTCCCTCCGCTCCGGAGATCCCAAGTGGAGCTTGA 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 311 GTGCTTACCACTCTGCTGTGACCCGAGCTTACCAAGGCTCTTGAAGGCAATGGA 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 652 GGTGCTCCAGCTGCTGCAGGCAACAGTCCAGAGGTGACCCCGGCAGAGTC 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 371 GCTGCCACCACTGTGAGAGGAGGGGTGCAGTGGAGGC 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 95
CF911441 575 bp mRNA linear EST 05-NOV-2003
LOCUS A0615D04-5 NIH Mouse Hematopoietic Stem Cell (Lin-/c-Klt-/Sca-1+)
DEFINITION cDNA library (long 1) Mus musculus cDNA clone NIH:A0615D04
IMAGE:30749031 5', mRNA sequence.
ACCESSION CF911441
VERSION CF911441.1 GI:38182645
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 575)
Piao, Y., Ko, N.-T., Lim, M.-K. and Ko, M.-S.-H.
Construction of long-transcript enriched cDNA libraries from
subprogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE 21429098
PUBMED 11544199
JOURNAL Dawood B. Dudekula
CONTACT: Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: A0615 row: D column: 04

Seq primer: M13 Reverse
High quality sequence stop: 575
POLYA=No.

FEATURES

source

Location/Qualifiers
1. 575
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6Ncr"
/db_xref="taxon:10090"
/clone="NIA:A0615D04 IMAGE:30749031"
/dev_stage="Age ~10 weeks old"
/lab_host="D10B"
/clone_11b="NIA Mouse Hematopoietic Stem Cell
(lin-/C-Kit-/Scd-1+) cDNA library (long 1)"
/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
Site 2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://lgen.gsc.nia.nih.gov/cDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
RNAs were obtained from Drs. Dennis Taub, Dan Longo
(National Institute on Aging, USA), Jonathan Keller
(National Cancer Institute, USA). Double-stranded cDNAs
were synthesized with an Oligo(dT) primer (Invitrogen):
5'-pGACTGTTTACATGATCGACGCGCGCTTTTCTTTT-3' from
1.1 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker Lr-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer SalI-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 2.2 Kb. The library was
constructed by Yulan Piao."

ORIGIN

Query Match 3.8%; Score 58.2; DB 7; Length 575;
Best Local Similarity 58.3%; Pred. No. 0.022;
Matches 102; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGCGGACGCGGAGCTCATCTGCTGACGCGCTGCTC 591
DB 173 AGGATTACTGTGAGGTGCGACAGGCGGGAATTATTTCTGTGCGACACTGCCCC 232
QY 592 GGGCCTTCCACTGCGCTGCTGCTCCCTCCGCGGAGATCCCAAGTGGAGCTTGA 651
DB 233 GTGCTTACCACTGCTGCTGCTTGAACCCGAGCTTACCGGGCTCTCTGAGGGCAAGTGA 292
QY 652 GGTGCTCAGCTGCTGCTGCGACGCAAGTCCAGAGAGTGCAGCCCGGGGAGAGGA 706
DB 293 GCTGCCCCCACTGTGAGAGAGAGGGGTGAGTGGAGGCCAAGAGAGAGAGGA 347

RESULT 96 690 bp mRNA linear EST 15-JUL-2003
LOCUS BU703948
DEFINITION UI-M-F00-bcr-e-16-0-UI-r1 NIH-BMAP_F00 Mus musculus cDNA clone
IMAGE:6406263 5', mRNA sequence.
ACCESSION BU703948
VERSION BU703948.1 GI:23631529
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 690)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

The following repetitive elements were found in this cDNA
sequence: 1-42, >(CAG)n\$Simple_repeat 647-650,
>(GGA)n\$Simple_repeat
Seq primer: pYX-5.

FEATURES

source

Location/Qualifiers
1. 690
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6406263"
/issue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="D10B (T1 phage resistant)"
/clone_11b="NIH BMAP_F00"
/note="Organ: Brain; Vector: pYX-Anc; Site 1: BclI;
Site 2: NotI; The library was constructed according
to Benton, Lennan and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a NotI site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with BclI adaptor, digested with NotI and then cloned
directionally into pYX-Anc vector. The library tag
sequence located between the NotI site and the polyA tail
is TAGAGAGCC. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN

Query Match 3.8%; Score 58.2; DB 5; Length 690;
Best Local Similarity 58.3%; Pred. No. 0.022;
Matches 102; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 532 AGGACGAGTGTGCGCTGTGCGGACGCGGAGCTCATCTGCTGACGCGCTGCTC 591
DB 513 AGGATTACTGTGAGGTGCGACAGGCGGGAATTATTTCTGTGCGACACTGCCCC 572
QY 592 GGGCCTTCCACTGCGCTGCTGCTCCCTCCGCGGAGATCCCAAGTGGAGCTTGA 651
DB 573 GTGCTTACCACTGCTGCTGCTTGAACCCGAGCTTACCGGGCTCTCTGAGGGCAAGTGA 632
QY 652 GGTGCTCAGCTGCTGCTGCGACGCAAGTCCAGAGAGTGCAGCCCGGGGAGAGGA 706
DB 633 GCTGCCCCCACTGTGAGAGAGAGGGGTGAGTGGAGGCCAAGAGAGAGAGGA 687

RESULT 97 2253 bp mRNA linear HTC 03-APR-2004
LOCUS AK045449
DEFINITION AK045449 full-length enriched library, clone:B23020M16 product:MI-2
autantigen 240 kDa protein (fragment) homolog [Homo sapiens], full
insert sequence.
ACCESSION AK045449
VERSION AK045449.1 GI:26337378
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE AUTHORS 5

REFERENCE
AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE AUTHORS 6 (bases 1 to 2253)

REFERENCE
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuta, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Iino, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurinara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takekida, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Yamanaka, M., and Hayashizaki, Y.
TITLE Direct Submision
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
Location/Qualifiers
1. .2253
/organism="Mus musculus"
/mol_type="mRNA"

FEATURES
source

/strain="C57BL/6J"
/db_xref="PANTOM:DB:B230201M16"
/db_xref="taxon:10090"
/clone="B230201M16"
/sex="male"
/tissue_type="corpora quadrigemina"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
129_->2252
/note="unmaned protein product, Mi-2 autoantigen 240 kDa protein (fragment homolog [Homo sapiens] (PIR:J13858), evidence: PASTY, 96.8%id, 98.4%length, match=1578)
putative"
/protein_id="BAC33375.1"
/cdon_start=1
/db_xref="GI:26337379"
/translation="WRRRTKRYSGAGRGHDSRDRSHSPSCHLFPPEPPPLPPPPDKDDIRLPSALGVKRRKPKQKQKRRKRRKRLDSEEFSEDEYREKESGQSEYDGTGRKRRKRRKREKRRKRRKRGEDGGHGVQVQKSAFTLLTWGLDVEDVSESDYHTLTWYKAFSOPMRPLAKQPKIPMSKMTITLAKMREFSANRPFKGSAAVAAAAAANAADVAASATPIAPSGPPALPPPAPEIQPPPIRRAKTREGKGGHRRNKSFPVPPGRKKLRKQWAPLKITGLGCKRKAGSAPFQSEDEPEABESDLDSGVSHSASGMPDGVRAKRLKRGPRKRVAGCAVTSBEGVDYETTHODYCEVQOQGEITLCTCPRAHVLCLDELDAAPGKSCPCBCEGVQWAKKEEBEYEEBEGEKEEEDHMEYCRVCDXGBELCDACISYHICLNPPLPDIENGMLCPCTCPVKGKRYOKILHMRGEPVAVPAPQADNPVPPRPLQGSEREFVYMGGLSVHGSWAKELOLEIFHLVWYRNYORNDDEPPLDYGSGEDGSKKVDPHYAMBEKTYRREGIKPEMMTVRIHSHSDKKNNHYIVKMDLPYDQSTWEDENNIPETDHHKQSTWREHLMGEDPAPPK"

ORIGIN
Query Match 3.8%; Score 58.2; DB 3; Length 2253;
Best local similarity 58.3%; Pred. No. 0.026;
Matches 102; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 532 AGCAGAGTGCCTGCTGTCGCGAGCGCGGAGCTCATCTGCTGTGACGCGTGCCTC 591
DB 1300 AGGATTACTGTGAGGTGTCGACAGCGCGGGAATTATCTGTGCCACACCTGCCCCC 1359

QY 592 GGGCTTCACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 651
DB 1360 GTGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1419

QY 652 GGTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 706
DB 1420 GCTGCCCACTGTGAGAGGAGGGGTGCTGAGAGGCCCAAGAGGAGGAGA 1474

RESULT 98
CNS00720
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
CNS00720 932 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR14B09 of RPL1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL066742
AL066742.1 GI:4945205
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 932)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Ooeagawa and Aaron Mammoser in Pieter de Jong's laboratory in the department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; on bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1..932
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14809"
/clone_lib="RPCL-98"
/note="end : T7"

ORIGIN

Query Match 3.8%; Score 58; DB 9; Length 932;
Best Local Similarity 31.3%; Pred. No. 0.026;
Matches 102; Conservative 77; Mismatches 147; Indels 0; Gaps 0;
QY CGCTGCAGGTCTGACCTCTCGCCCTGACACCCCTACTGTGTGAGTCTGAGGCTC 918
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 531
CGGKGTGKGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 590
QY AGCAGAACCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 918
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 591
YTBKGGGCGGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSG 650
QY GGTGTACTACTGCGCGCTGCTCTTCACTGCGCTGCTGCTGCTGCTGCTGCTGCT 1038
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 651
GCCGCCSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSG 710
QY GGGCGGAGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1098
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 711
GSCGCGGCGGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSG 770
QY AGGGGCTGTGCGCCCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCG 1158
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 771
CGSGCGGCGGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSG 830
QY 1159 CCAGTACAGACCCGCTCTGACAGG 1184
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 831
CCSSGSGCGCCSSGCGCGCGSSSS 856

RESULT 99
AG057392 997 bp DNA linear GSS 02-NOV-2001
LOCUS Pan troy1odytes DNA, clone: PTB-043P11.R, genomic survey sequence.
DEFINITION AG057392
ACCESSION AG057392
VERSION AG057392.1 GI:16594851
KEYWORDS GSS.
SOURCE Pan troy1odytes (chimpanzee)
ORGANISM Pan troy1odytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.

REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Toto, Y., Watanabe, H., and Sakaki, Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 997)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Toto, Y., Watanabe, H., and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenhiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpbes@gsc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>).

COMMENT

Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: M13Rev
LIBRARY
Vector : PKS145
R.site 1 : SacI
R.site 2 : SacI.

FEATURES

source
1..997
/organism="Pan troy1odytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-043P11.R"
/sex="male"
/cell_type="Lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN

Query Match 3.8%; Score 58; DB 9; Length 997;
Best Local Similarity 44.6%; Pred. No. 0.026;
Matches 229; Conservative 0; Mismatches 285; Indels 0; Gaps 0;
QY 696 CGGCGAG 755
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 121
CGGCGAG 180
QY 756 AGGTGCGGAG 815
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 181
GCGCGCGAG 240
QY 816 ACTTGTGTACAG 875
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 241
GCGCGCGAG 300
QY 876 TCTTGGGCTTGCAG 935
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 301
GCGCGCGAG 360
QY 936 GGTGCGGCTTGCAG 995
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 361
GCGCGCGAG 420
QY 996 GCTGCTTGCAG 1055
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 421
GCGCGCGAG 480
QY 1056 CGTGTGAGATCTGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1115
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 481
CCCGCGCGAG 540
QY 1116 AGCGCGCGAG 1175
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 541
CGCGCGCGAG 600
QY 1176 CTGCAAG 1209
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 601
CGGCGCGAG 634

RESULT 100

BO953925 1038 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT 8803230 lusek1 sciatic nerve Homo sapiens cDNA clone
DEFINITION IMAGE:6198951 5', mRNA sequence.
ACCESSION BO953925
VERSION BO953925.1 GI:22369403
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 1 (bases 1 to 1038)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
 Plate: LIML3611 row: 9 column: 16
 High quality sequence stop: 214.
 Location/Qualifiers

FEATURES

source
 1..1038
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6198951"
 /sex="male"
 /tissue_type="sciatic nerve"
 /dev_stage="adult, 70 yr"
 /lab_host="DH108"
 /clone_lib="Lupski, sciatic nerve"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1: NotI; Site 2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCAGCGCTCCG-3' and 5'-GACTAGCTTACATCGGAGGCGCCGCTT(15)-3'. Size selected > 1 kb for average insert length 1.87 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

ORIGIN

Query Match 3.7%; Score 57.8; DB 5; Length 1038;
 Best Local Similarity 37.9%; Pred. No. 0.029;

Matches 307; Conservative 0; Mismatches 498; Indels 6; Gaps 2;

```

QY 290 CCGGGTTTCTTCCCAATAGGATGGCCCGGGGGGTCTCTTCGAGACCAATGATGATG 349
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 CCAAGCTTAGTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 350 GGGAAAGTGTGTCAAGGCAAAATTTCAAGCCCTGCGACATGGAGCAGAGCAAGACT 409
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 GNGNNGGGGGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNG 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 410 GGGAGTTCAGTACCAAGAG-ATGCTGCTGGGGGAGCTTTTGGAAAGAGTGGCTC 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 GGGGGGCGCGGNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 374
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 469 TCAGAGAGGTGTGCAACCCAGCCAGTCTGATGGGCGTCTTGTGCTGTGCCAGAAGA 528
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 CANNGANNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 529 ATGAGAGCGAGTGTGCTGTGTGTGAGGACGCGGAGAGCTCATCTGCTGTGACGCTGCC 588
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 CCCCCCGGCGCGGNGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 494
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 589 CTGGGGCTTTTCCACTGTGGCTGTGCTGTGCTGTGCGGAGATCTCCCAATGGGACTT 648
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 CCNGGGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 554
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 649 GAGAGTCTTCAGAGTCTGTGAGGCAAGTCCAGAGGTGAGAGAGAGAGAGAGAGC 708
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 555 GCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 709 CCGGGCCCCAGAGCAACCCGTGAGAGACCCGCTCCCGGGGGCTTAAGTGGCGGGAG 768
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 615 CCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 674
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 769 AGAGGTAAAGAGTCCACTTGGGGAACCCCTTACCGGCAATGACAGACTTTGTATA 828
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 GCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 734
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 829 AGACCTGGCG-----GTCGCGCTTCTGTGACAGCCCGCTGCCAGGTCTGAATCTCTGAGC 883
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 735 NCCNNNGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 794
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 884 CCTGCACCCCTTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 943
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 795 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 854
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 944 TTGCGGGGTGTGCGGAGATGTATCGGAGCTGTGGGTACTCACTGCGCGCTGCTT 1003
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 855 NCGCGCGGNNCCCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 914
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1004 CCACTGGCGCTGCACTTCCAGCGGACCTCCCGGACCGGAGACGGGCTGCTGACG 1063
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 915 GCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 974
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1064 ATCTGTCTCAGGAGACGTGACCCCGAGCCCTT 1094
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 975 CTCGCGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1005
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: February 9, 2005, 17:40:05
 Job time : 4565 secs